DOCUMENT RESUME

ED 061 722 EM 009 705

AUTHOR	Hedl, John Joseph, Jr.
TITLE	An Evaluation of a Computer-Based Intelligence Test.
INSTITUTION	Florida State Univ., Tallahassee. Computer-Assisted Instruction Center.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
PUB DATE	1 Jun 71
NOTE	78p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Anxiety; *Computer Programs; Group Intelligence Testing; Intelligence Quotient; *Intelligence Tests; Measurement Techniques; Psychological Evaluation; Psychological Tests; Student Evaluation; *Testing
IDENTIFIERS	SIT; *Slosson Intelligence Test

ABSTRACT

A study demonstrated the feasibility, reliability and validity of a computer-administered intelligence test. The computer-based test (Slosson Intelligence Test) was compared with a conventional examiner-administered test given to the same college students within the same week. In addition it was evaluated as a predictor of college grade point average. It scored high in both tests, being the best predictor of future college grade point average. The computer testing experience, however, led to higher levels of anxiety than normal examiner testing sessions. The attitudinal and anxiety data presented indicate that there might be a complex interaction between the nature of the test and the quality of the computer programing procedures that in turn determines the affective response. (RB)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

CAI CENTER

ED 061722

AN EVALUATION OF A COMPUTER-BASED INTELLIGENCE TEST

John Joseph Hedl, Jr.

Tech Report No. 21
June 1, 1971

Project NR 154-280
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Arlington, Virginia
Contract No. N00014-68-A-0494

Approved for public release; distribution unlimited.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

EM 009 705

FLORIDA STATE UNIVERSITY

AN EVALUATION OF A COMPUTER-BASED INTELLIGENCE TEST

John Joseph Hedl, Jr.

Tech Report No. 21
June 1, 1971

Project NR 154-230
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Arlington, Virginia
Contract No. N00014-68-A-0494

Approved for public release; distribution unlimited.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Security Classification

DOCUMENT CONTROL DATA–R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Florida State University Computer-Assisted Instruction Tallahassee, Florida 32306	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

AN EVALUATION OF A COMPUTER-BASED INTELLIGENCE TEST

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

John Joseph Hedl, Jr.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1, 1971	72	60

8a. CONTRACT OR GRANT NO. N00014-68-A-0494 b. PROJECT NO. NR 154-280 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Tech Report No. 21 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Personnel & Training Research Program Office of Naval Research Arlington, Virginia

13. ABSTRACT

The general purpose of the present study was to investigate a computer application to the area of individualized intelligence testing. Primary emphasis was placed on questions relating to the feasibility, reliability, and validity of a computer-based administration and scoring program for the Slosson Intelligence Test (Slosson, 1963). Specifically, the evaluation focused on concurrent validity relationships with the Wechsler Adult Intelligence Scale (Wechsler, 1955). In addition, predictive validity in terms of college grade point average was compared in terms of forecasting efficiency. Certain affective variables (student attitudes and anxiety) were studied to determine their relationship to the computer-based and traditional examiner test administration procedures. Examiner ratings of student testing behaviors were related to the attitudes and anxiety variables.

The computer-based Slosson Intelligence Test (CB-SIT) was designed to operate with an IBM 1500 Instructional System (IBM, 1967). The test items were individually presented on a cathode ray tube. Students entered their answers for immediate computer evaluation. The answer analysis algorithms were based upon a key word/phrase dictionary for each item which was developed from previous test item protocols. If the student was scored as either correct or incorrect, the program continued to the next item. If the answer was scored

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

DD FORM 1473
1 NOV 65
S/N 0101-6811

(BACK)

4

Security Classification
A-31409

as partially correct, the student was instructed by the computer to more fully explain his response.

Forty-eight undergraduate students (24 males and 24 females) were individually tested with the WAIS, SIT, and CB-SIT. A latin square design was employed to counterbalance potential order and sequence effects of the three different intelligence test administrations. In addition, each testing session included an administration of an attitude scale and anxiety scale both before and after each testing period. A time period of approximately one week separated the different test administrations.

Within the context of the latin square design, the CB-SIT was shown to correlate .75 with a traditional examiner administration of the SIT. More importantly, equivalent concurrent validity relationships with the WAIS were shown as follows: the CB-SIT correlated .54 with the WAIS and the SIT correlated .52 with the WAIS with a college population. In terms of predictive validity, the best predictor of college grade point average in terms of R^2 was the CB-SIT.

In terms of the affective measures, the computer testing experience lead to higher levels of state anxiety in comparison to either of the two examiner testing sessions. The student attitude results were fairly consistent with the anxiety results. Generally, attitudes toward examiner testing increased for the pre to post measurement periods, whereas attitudes toward the computer testing procedures decreased over the same measurement periods.

Examiner ratings of test behavior were found to be significantly related to the performance of the **S**s. In addition, these ratings were also shown to be positively related to the student attitude measures and negatively related to the state anxiety scores.

The results of the present investigation demonstrated the feasibility and validity of an interactive computer approach to the assessment of intelligence. In addition, the attitudinal and anxiety data indicated that there might be a complex interaction between the nature of the test and the quality of the computer programming procedures that in turn determines the affective nature of automated testing procedures.

TABLE OF CONTENTS

ABSTRACT vi
ACKNOWLEDGEMENTS viii
LIST OF TABLES ix
LIST OF ILLUSTRATIONS xi

Chapter

I. INTRODUCTION 1

II. REVIEW OF LITERATURE 3
- Technological Applications in Intelligence Testing
- Computer Testing Applications Using Language Processing
- Anxiety and Computer-Based Evaluation
- Statement of the Problem

III. METHOD 8
- Description of Computer-Based Slosson Intelligence Test Development
- Evaluation Instruments
  - Slosson Intelligence Test
  - Wechsler Adult Intelligence Scale
  - State-Trait Anxiety Inventory
  - Attitude Inventory
  - Examiner Rating Scale
- Subjects
- Apparatus
- Experimental Design
  - Pretask Period
  - Task Period
  - Posttask Period

IV. RESULTS 15
- Evaluation of Test Administration Procedures
- Reliability Analysis of the CB-SIT and SIT
- Concurrent Validity Evaluation
- Predictive Validity of the WAIS, SIT, and CB-SIT in Relation to College Grade Point Average
- Evaluation of State Anxiety
- Evaluation of Student Attitudes

Developmental Problems and Suggestions for Future Revisions
Test Presentation Strategy Analysis and Modifications
Scoring Analysis and Modifications

V. DISCUSSION 35

VI. SUMMARY 40

REFERENCES 42

APPENDICES
Appendix A 49
Appendix B 53
Appendix C 57
Appendix D 61

VITA 63

ABSTRACT
AN EVALUATION OF A COMPUTER-BASED INTELLIGENCE TEST

John Joseph Hedl, Jr.
The Florida State University

The general purpose of the present study was to investigate a computer application to the area of individualized intelligence testing. Primary emphasis was placed on questions relating to the feasibility, reliability, and validity of a computer-based administration and scoring program for the Slosson Intelligence Test (Slosson, 1963). Specifically, the evaluation focused on concurrent validity relationships with the Wechsler Adult Intelligence Scale (Wechsler, 1955). In addition, predictive validity in terms of college grade point average was compared in terms of forecasting efficiency. Certain affective variables (student attitudes and anxiety) were studied to determine their relationship to the computer-based and traditional examiner test administration procedures. Examiner ratings of student testing behaviors were related to the attitudes and anxiety variables.

The computer-based Slosson Intelligence Test (CB-SIT) was designed to operate with an IBM 1500 Instructional System (IBM, 1967). The test items were individually presented on a cathode ray tube. Students entered their answers for immediate computer evaluation. The answer analysis algorithms were based upon a key word/phrase dictionary for each item which was developed from previous test item protocols. If the student was scored as either correct or incorrect, the program continued to the next item. If the answer was scored as partially correct, the student was instructed by the computer to more fully explain his response.

Forty-eight undergraduate students (24 males and 24 females) were individually tested with the WAIS, SIT, and CB-SIT. A latin square design was employed to counterbalance potential order and sequence effects of the three different intelligence test administrations. In addition, each testing session included an administration of an attitude scale and anxiety scale both before and after each testing period. A time period of approximately one week separated the different test administrations.

Within the context of the latin square design, the CB-SIT was shown to correlate .75 with a traditional examiner administration of the SIT. More importantly, equivalent concurrent validity relationships with the WAIS were shown as follows: the CB-SIT correlated .54 with the WAIS and the SIT correlated .52 with the WAIS with a college population. In terms of predictive validity, the best predictor of college grade point average in terms of R^2 was the CB-SIT.

In terms of the affective measures, the computer testing experience lead to higher levels of state anxiety in comparison to either of the two examiner testing sessions. The student attitude results were fairly consistent with the anxiety results. Generally, attitudes toward examiner testing increased for the pre to post measurement periods, whereas attitudes toward the computer testing procedures decreased over the same measurement periods.

Examiner ratings of test behavior were found to be significantly related to the performance of the Ss. In addition, these ratings were also shown to be positively related to the student attitude measures and negatively related to the state anxiety scores.

The results of the present investigation demonstrated the feasibility and validity of an interactive computer approach to the assessment of intelligence. In addition, the attitudinal and anxiety data indicated that there might be a complex interaction between the nature of the test and the quality of the computer programming procedures that in turn determines the affective nature of automated testing procedures.

ACKNOWLEDGEMENTS

The author is extremely grateful to Dr. Duncan N. Hansen, who directed this dissertation and generously contributed his time, patience, and invaluable critical insights to this research. Dr. Hansen sustained and supported me both intellectually and personally throughout my graduate career at Florida State University. My debt to Dr. Hansen is a tremendous one which will never be forgotten. The author also acknowledges the timely advice and suggestions of my committee members.

The author would like to thank the entire staff at the CAI Center for their understanding, friendship, and assistance during the past three years. The author is indebted to Mrs. Dot Carr for her excellent preparation and care in the typing of this dissertation.

The author would like to express a special thanks to his friend and colleague, Dr. Harold F. O'Neil, Jr., for his emotional support and confidence during the critical periods in the preparation of this document. His technical and editorial assistance has been invaluable.

The author would also like to thank all of the graduate students who assisted in the implementation of this investigation and to those friends who provided the necessary encouragement.

This dissertation is dedicated to my parents who supported me in countless ways during my educational career.

LIST OF TABLES

1 Means and Standard Deviations of WAIS, SIT, and CB-SIT IQ Scores by Entire Sample and for Experimental Groups 15

2 Analysis of Variance for Tests and Test Administration Procedures 16

3 Modified Split-Half Reliability Coefficients for the CB-SIT and SIT by Entire Sample and for Experimental Groups 17

4 Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Entire Sample (N=48) 18

5 Coefficient of Correlation Between CB-SIT, SIT, and WAIS Subtests for Entire Sample (N=48) 19

6 Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Group I (n = 16) 20

7 Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Group II (n=16) 21

8 Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Group III (n=16) 21

9 Multiple Regression Analysis with GPA As the Criterion (N=48) 22

10 Kuder-Richardson 20 Reliability Estimates for the STAI A-State Measures . . 24

11 Means and Standard Deviations of STAI A-State Scores by Intelligence Test for Experimental Groups and Entire Sample 24

12 Analysis of Variance of STAI A-State Scores by Intelligence Tests, Test Sessions, and Pre-Post Measurement Periods 25

13 Means and Standard Deviations of STAI A-State Scores as a Function of Trait Anxiety, Sex, and Intelligence Tests 26

14 Kuder-Richardson 20 Reliability Estimates for the Attitude Scale Measures . . 26

15 Means and Standard Deviations of Attitude Scale Scores by Intelligence Test for Experimental Groups and Entire Sample 27

16 Analysis of Variance of Attitude Scale Scores by Intelligence Tests, Test Sessions, and Pre-Post Measurements Periods 28

17 Coefficients of Correlation for Total WAIS Examiner Rating Scale Scores . . . 29

18 Coefficients of Correlation for Total SIT Examiner Rating Scale Scores 29

19 Analysis of Variance for Examiners 33

20 Comparison of Computer and Examiner Scoring Methods for Initial Field Trial and Present Study 34

LIST OF ILLUSTRATIONS

1 Experimental design with number of Ss 13

2 Item 24-9 from the SIT 31

CHAPTER I
INTRODUCTION

A controversial issue in psychological assessment concerns the extent to which technology, particularly the computer, is to be used in future diagnostic evaluations. The importance of this issue is reflected in the broadening body of literature concerned with the efficacies of computer testing (Archambault, 1970; Elwood, 1969; Finney, 1967; Fowler, 1964, 1969; Gorham, 1967; Piotrowski, 1964; Veldman, 1967). Advocates for the utilization of the computer have cited the logistic and cost-effective benefits to be derived from computer-based evaluation. The shortage of professional personnel to administer, score, and interpret psychological test data has been well-documented (Arnhoff, 1968; Boneau, 1968a; Boneau, 1968b). With current projections revealing the inadequacies of present training programs in producing the necessary competent manpower, computer technology, it is argued, offers a viable alternative for solving these logistic problems.

In addition to these considerations, computer-based evaluation may have important reliability and validity implications. Computer-based administration of psychological tests should increase the reliability and validity of the test information. Since the computer is conceptually objective and neutral, its use to administer tests should eliminate possible human biases resulting from the typical dyadic interaction between examiner and student or patient. The reduction of these affective error variance components should lead to increased reliability of the tests (Cronbach, 1960).

Automation of test administration may also increase the validity of personal student information. Smith (1963) points to a "Confession Machine Effect" which appears to enhance the data acquisition in particular content areas such as the subject's personal experience or his perceived personality characteristics. Results of an investigation (Lawler, 1968) involving testing of the concepts of reproduction, birth control, and social morals concerning sex, bear directly to this point. The college students found the anonymous role within computer testing to be an enjoyable, non-threatening experience. Cogswell and Estavan (1965) have also reported similar findings on the apparent confidentiality of the computer interview.

Computer-based scoring and analysis profiling of test scores has perhaps been the easiest and most widely implemented computer application in psychological evaluation. Such procedures had previously been developed for use with aptitude and achievement tests and little difficulty has been encountered in adapting such procedures to psychological tests of similar nature. The most important benefit to be derived from computer scoring procedures is the complete objectivity and accuracy of the process. Given the same data, the computer always produces the same results. If errors are present in the scoring program, they will be consistent and easier to amend.

The final phase of the psychodiagnostic process consists of data interpretation and prediction. A debate has been raging during the past few years over the relative merits of actuarial methods versus clinical methods for interpretation and prediction purposes. Meehl (1956) and later Marks and Seeman (1963) have indicated that statistical decision-making may be as insightful and as reliable as the intuitive interpretative processes applied by the clinician. However, other psychologists (Holt,

1958; Holtzman, 1960; McArthur, 1956) feel that the computer will only serve a minor role for interpretation due to the judgmental nature of the interpretation process.

The general purpose of the present study was to investigate a computer application to the area of individualized intelligence testing. Primary emphasis was placed on questions relating to the feasibility, reliability, and validity of a computer-based administration and scoring program for the Slosson Intelligence Test (Slosson, 1963). Specifically, the evaluation focused on concurrent validity relationships with the Wechsler Adult Intelligence Scale (Wechsler, 1955). In addition, predictive validity was compared in terms of forecasting efficiency. Certain affective variables (student attitudes and anxiety) were studied to determine their relationships to the computer-based and traditional intelligence test administration procedures. Examiner ratings of student testing behaviors were related to the attitude and anxiety variables.

CHAPTER II
REVIEW OF LITERATURE

The primary objectives of this study were to investigate the feasibility, reliability, and validity of a computer-based intelligence test. In addition, certain affective variables were explored to determine their relationship to the computer and examiner test administration procedures. In order to provide the background for this investigation, this review will cover three areas: (1) technological applications in intelligence testing; (2) computer testing applications with natural language processing; and (3) anxiety and computer-based evaluation. The final section presents an explicit statement of the goals for the present investigation.

Technological Applications in Intelligence Testing

Although individual intelligence testing remains a major time-consuming endeavor for school and clinical psychologists, no computer programs have been developed to administer, score, or interpret the test data. However, recent findings in this area of evaluation have demonstrated the feasibility of a technological approach and paves the way for further research and developmental efforts.

As an illustration of the usefulness of automated systems in providing psychological testing services to patients, Elwood (1969) has developed an automated testing booth programmed to administer the Wechsler Adult Intelligence Scale (Wechsler, 1955). The testing apparatus is capable of automatically administering all verbal and performance subtests of the Wechsler Adult Intelligence Scale (WAIS). The system records the patient's responses to the individual test items, scores the Digit Symbol and Digit Span subtests, and provides several temporal measures for item responses.

Griffith and Elwood (1969) have reported favorable test-retest reliability coefficients for the automated WAIS administration (Verbal IQ, r = .99, Performance IQ, r = 97, and Full Scale IQ, r = .99). In another study, Elwood and Griffith (1971) reported test-retest coefficients of .97, .95, and .98 for VIQ's, PIQ's, and FS-IQ's, respectively. Orr (1969) compared the automated WAIS with a traditional administration of the WAIS. Concurrent validity coefficients between the two administration methods were as follows: VIQ's, r - .94; for PIQ's, r = .82. The correlation between FS-IQ's was 9.3. The main problem concerning the comparability of the two methods seems to lie primarily in the performance section.

The initial WAIS data are quite encouraging and demonstrate the feasibility of automated individual intelligence test administrations. However, the system only provides scoring capabilities for two of the 11 subtests. In addition, the test protocols must still be scored and interpreted by a psychologist upon completion of the testing session (average time approximately 25 minutes per test). Thus, the next obvious step would be to automate both the test administration and scoring for an individual intelligence test.

We shall now turn to a review of other computer applications in psychological assessment. Special emphasis will be placed on those applications which involve the computer in natural language processing problems.

Computer Testing Applications Using Language Processing

Although employment of computers to calculate test scores and to carry out statistical summaries and analyses of test data has been common for many years, the extension of these methods to encompass more than multiple-choice questionnaires has been a relatively recent development. The problems of interpreting natural language data are far more complex than those involved in the translation of score profiles such as has been performed with the Minnesota Multiphasic Personality Inventory (Rome, Swenson, Mataya, McCarthy, Pearson, & Keating, 1962), Gough's California Psychological Inventory (Finney, 1967), and for Cattel's 16 Personality Factors Inventory (Eber, 1964).

As indicated above, relatively efficient procedures are available for producing readable summaries from a limited range of psychological instruments. However, the potential applications are far more significant for the construction of testing programs to process natural language, even to a limited extent.

Standing virtually alone in the field of computer-based content analysis is the General Inquirer System (Stone, Dunphy, Smith, & Ogilvie, 1966.) Using special "dictionaries" of words precategorized for specific research purposes, the system automatically tallies frequencies of category usage for a body of text material. The materials which have been analyzed range form suicide notes (Stone, et al., 1962) to Thematic Apperception Test narratives (Smith, 1968).

Most applications of the General Inquirer have ignored the problems of syntax. Goldberg (1966), for instance, applied the system to sentence completions with some success. Other researchers in the field of automated content analysis have evaded syntax problems by restricting the responses of the subject in one manner or another. In developing a computer-based system for scoring responses to the Holtzman Ink-blot Test, Gorham (1967) restricted subjects to the use of six words for each blot. Even with this restriction, the correlations between hand and computer scoring equalled or exceeded interscorer reliability for the computer scoring for 15 of the 17 variables.

Peck and Veldman (1961) of the University of Texas have been developing a computer-based system for scoring responses to a sentence completion test. The problems of syntax were reduced due to the restriction on the subject to use a single word in responding to each sentence stem. The most recent system produces 40 scores from a 36-item form and employs a complex word-root data reduction system (Veldman, 1967). This tailored inquiry method offers many of the benefits of a traditional interview, but it utilizes a relatively constrained stimuli, that is, sentence completion responses are a limited subset of the responses found in word definitions.

Recently, Archambault (1970) developed a computerized program to score verbal responses to three of the seven subtests of the Torrance Tests of Creative Thinking. The subtests considered were the Ask and Guess subtests (Activities 1, 2, and 3) in which subjects ask questions about a drawing and make guesses about the causes and consequences of a pictured event. Subject responses to each of these subtests were scored for Fluency, Flexibility, and Originality.

For each of these categories a dictionary of entries was constructed by analyzing the model responses given by Torrance for key words and phrases in *Roget's International Thesaurus* (1962) and *Soule's Dictionary of English Synonyms* (1966).

The test was administered in traditional fashion and the student responses were keypunched on standard IBM cards, one response to a card. These responses were then analyzed in a batch process mode. A word/phrase lookup procedure was performed to determine the frequency of categories which were used.

Archambault's data indicated that creativity, as defined by Torrance, was judged accurately by a computer. The syntax problems were reduced by only analyzing the frequency of word usage. However, this procedure produced significant correlations ranging from .52 to .99 between the computer and the pooled scores of four trained judges. It appears that the use of a computer to score open-ended responses to standardized test items is feasible and should be investigated.

In summary, the majority of these programs have used a dictionary approach for the analysis of the test responses. Many of the programs have evaded or ignored entirely the problems of syntax by restricting the subject in one way or another. However, significant relationships were found between computer scoring and clinical scoring. Perhaps in computerized scoring of test responses, the use of syntax relationships in the decision process concerning the nature of the test response may not be absolutely essential. The program designed to administer and score the Slosson Intelligence Test uses a dictionary approach, which by design, minimizes the analysis of syntactical relationships.

Another area of interest in computer testing concerns the nature of the man-machine interactive experience. The following section briefly reviews this area with an emphasis on anxiety and computer-based evaluation.

Anxiety and Computer-Based Evaluation

Many investigations have asserted the improved objectiveness and confidentiality of the computer testing experience (Cogswell and Estavan, 1965; Smith, 1963; Smith, 1968). However, only limited evidence from CAI research dealing with Trait-State Anxiety Theory (Spielberger, Lushene & McAdoo, 1971) exists that supports this assertion. In two studies (O'Neil, Spielberger, & Hansen, 1969; O'Neil, Hansen, & Spielberger, 1969) the CAI learning experience did not seem to differentially effect state anxiety responses for high and low trait anxiety Ss, although there was a significantly higher state response by high trait anxiety Ss. An analysis of the CAI situation revealed a possible explanation for the absence of any relationship between trait anxiety and differential increases in state anxiety in this CAI setting. In the CAI task, the computer did not evaluate the adequacy of the S's performance relative to others, and therefore, did not pose a threat to self-esteem. These two studies, because they did not find differential shifts on A-State results for low and high trait-anxious Ss, lend indirect evidence for the supposed impersonal nature of a computer task.

More direct evidence for the non-threatening nature of a computer-based evaluation comes from a study by Gallagher (1970). He investigated the relationship of instructional treatments and learner characteristics in a computer-managed instruction course. Computer evaluation and instructor evaluation of term projects resulted in some rather interesting findings. Trait anxiety scores were related to performance ($r = -.51$) in the instructor evaluated group, but were not in the computer evaluated group ($r = -.03$). If one assumes that the treatment group which emphasized human interaction

(instructor evaluated group) would result in a greater threat to the individual's self-esteem, then these results would be consistent with Trait-State Anxiety Theory. In addition, these results provide some evidence that the computer may be less threatening, and, therefore, may be more neutral in nature.

The present study further extended this research by direct comparison of anxiety measures within computer and examiner administrations of intelligence tests. In this respect, empirical evidence pertaining to the nature of a man-machine interaction was obtained.

The following section presents a more explicit statement of the research goals.

Statement of the Problem

It is apparent from the preceding research that computer technology has been successful for the automation of various psychological test instruments. However, the application of technology to individual intelligence testing has been a relatively recent and limited phenomena. Given that the assessment of intellectual functioning dominates a major portion of a psychologist's time commitment, and that intelligence test scores are still our best predictor of school success (Vernon, 1958), a reliable and valid computer-based approach has immediate applications for psychoeducational assessment services. The present investigation was concerned with the evaluation of a computer-based approach to individualized intelligence testing.

Specifically, the validity evaluation attempted to determine: (1) the relationship between a computer-based administration of the Slosson Intelligence Test and a traditional administration of the Slosson Intelligence Test (SIT), (2) the concurrent validity of these two Slosson test scores in comparison to the Wechsler Adult Intelligence Scale, and (3) the predictive efficiency of these intelligence test scores for the criterion of college grade point average (GPA).

In order to more fully explore the complex nature of this man-machine testing application, student attitudes and anxiety were assessed both before and after the different testing sessions. The major question to be answered here concerned the relationship of these variables to the varied test administration procedures. In addition, the nature of examiner ratings of students' test behavior was related to the attitudinal and anxiety responses of the Ss under the different testing conditions.

The following research hypotheses were evaluated in this investigation:

1) There would be significant relationships between computer-based SIT IQ scores, SIT IQ scores, and WAIS FS-IQ scores.

2) Significant prediction equations would be found for the three different IQ scores with GPA as the criterion measure.

The limited literature concerning anxiety and attitude measures and their relation to test administration method did not permit the derivation of explicit predictions for these measures. However, this investigation attempted to provide data to answer the following three exploratory questions:

1) What is the relationship between student testing attitudes and the method of intelligence testing administrations?

2) What is the relationship between levels of state anxiety and method of intelligence test administration?

3) What is the relationship between examiner ratings, student attitudes, and state anxiety scores?

CHAPTER III
METHOD

The present investigation evaluated a computer-based program designed to administer the SIT and to score the student answers on a real-time basis. Major emphasis was placed on validity questions by comparing the computer-based IQ scores with scores obtained from a traditional administration of the SIT and scores obtained on the same Ss on the WAIS. In addition, certain affective variables (attitude and anxiety) were studied to determine their relationship to the computer and examiner administration procedures. Examiner ratings were studied for possible relationships to the student self-report affective variables.

Description of Computer-Based Slosson Intelligence Test Development

The computer-based Slosson Intelligence Test (CB-SIT) has been designed to operate with an IBM 1500 Instructional System (IBM, 1967). The test items are presented individually on a cathode ray tube. Students then enter their answers for immediate computer evaluation. If the student is judged as totally correct or incorrect, the program continues to the next item. If he is judged as partially correct, the student is instructed to more fully explain his response.

In order to delimit the most effective scoring procedures for the program, it was necessary to determine the range and scope of expected student answers. To accomplish this, the vocabulary or word definition items of the SIT were administered via paper and pencil to 84 college students enrolled in an introductory habilitative science course at the Florida State University. The obtained responses were then scored by a skilled psychologist. Following these procedures, the correct answers were content analyzed and categorized according to similarities.

In essence, the first phase was to define the fundamental attributes that would determine a correct word definition. A specific example will clarify this procedure. The test word "plutocracy" contains two basic concepts in the standard meaning, that of government or control, and that pertaining to the rich or wealthy class of society. By using the sample data, and by searching *Roget's International Thesaurus* (1962) and *Soule's Dictionary of English Synonyms* (1966), two lists of word entries were constructed. Each dictionary set of entries contains those words which can be used to correctly describe each of the two concepts for the test item. The student, in order to obtain a correct answer, must match a word in each of the two concept categories. As can be seen, the correct answer to the test item "plutocracy" could be expressed in a number of different ways, each syntactically unique (e.g., government by the rich, where wealthy people rule a country, etc.). This procedure for keyword analysis is largely ignoring the syntactic relationships.

For other words, the procedure for analysis is similar. In certain instances, certain phrases are stored in the analysis buffers which must be matched to obtain a correct answer. The major problem encountered is that of synonomy (the expression of equivalences among words and sentences). The procedure of categorizing the test items into concepts or components has enabled efficient modifications to be implemented where necessary.

It should be obvious that the present answer sets do not contain all possible lexical elements for describing an answer. An heuristic principle was employed in the design of the answer dictionaires. For example, 90 percent of the sample answers to the test item "environment" were composed of the word "surroundings." This, then, defines the first analysis check point for the program. The most prevalent answers are analyzed first by the program before other lexical subroutines are used. In this respect, the developmental procedures focused on program efficiency.

For the first major field trial of the program, 25 students (14 males and 11 females) were obtained from an introductory psychology class at Florida State University during the Spring Quarter, 1970. Student test item protocols obtained from detailed computer performance records were typed and given to a skilled examiner for scoring. The examiner derived IQ scores from his individual evaluation of each student's performance data. This data was then compared to the computer-derived method of evaluation. The latter scoring method consisted of the author deriving scores based upon the program scoring decisions per item independently of any other evidence of student performance.

The correlation between the computer-derived IQ scores and the examiner-derived IQ scores was .83. The mean IQ for the examiner-derived IQ scores was 129 (sd=8.7). In comparison, the mean of the computer-derived IQ scores was 118 (sd=8.2). The small standard deviations were probably due to the homogeneous nature of the subject population.

The data strongly indicated the feasibility of an interactive computer-based method for the assessment of intelligence. On the other hand, the consistently lower computer-based IQ scores represent a serious problem. Item dictionary revisions based on this field trial data were implemented to amend these scoring discrepancies. This test was administered to determine the validity of the CB-SIT in relation to the SIT and the WAIS.

Evaluation Instruments

The instruments used to collect data to evaluate the CB-SIT were: (1) Slosson Intelligence Test (Slosson, 1963); (2) Wechsler Adult Intelligence Scale (Wechsler, 1955); (3) State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970); (4) an attitude scale modeled after the Semantic Differential (Osgood, Suci & Tannebaum, 1957); and (5) an examiner rating scale. In addition to these measures, college grade point averages were obtained for the Ss from the Office of the Registrar at Florida State University. A description of these instruments, the rationale for their use, and their psychometric properties follows.

Slosson Intelligence Test

The Slosson Intelligence Test (Slosson, 1963) was designed to be an individual screening instrument for the assessment of intelligence for both children and adults. The test has adapted many items from the Stanford-Binet, Form L-M (Terman and Merrill, 1960). The items are grouped according to difficulty, mixing content randomly. The major validity study using Stanford-Binet, Form L-M scores and Slosson Intelligence Test (SIT) scores showed concurrent validity coefficients of .90 to .98 with Ss (N=701) ranging in age from four years to 18 and above. In another study (Slosson, 1963),

a correlation of .92 was found between Binet and SIT scores for Ss ranging in age from four years to 19 (N=141).

Several small studies using the SIT have shown somewhat favorable results. Hammill (1967) reported a test-retest reliability of .97 for 155 Philadelphia school children. The coefficient of internal consistency for the total sample was .95. O'Keefe (1967) found a correlation of .75 between the Binet and SIT for 42 elementary students enrolled in a class for fast learners. Their ages ranged from 10 years, 2 months to 12 years, 1 month. DeLapa (1968) compared Binet and SIT scores for children with an age range of 8 years, 10 months to 11 years, 11 months. One group included 57 elementary students in four special education classes for the educable and another group included 60 students in regular elementary classes. The resulting correlation between Binet and SIT scores for the special students was .60 and for the regular class students, it was .90.

Mackinnon and Elliott (1969) administered the California Achievement Test (CAT), the Primary Mental Abilities Test (PMA), and the SIT to 127 children in 16 elementary special education classes. SIT scores correlated .62 with CAT scores and .67 with PMA scores. A correlation of .68 was found between the PMA scores and CAI scores. The authors concluded that the PMA may be more valid than the SIT for southern, elementary level, retarded children. The high attrition level of Ss in this study makes these results tenuous.

The only investigation using the Wechsler Adult Intelligence Scale (Wechsler, 1955) as criterion was one by Kaufman and Ivanoff (1969). They administered the Wechsler Adult Intelligence Scale (WAIS) and the SIT to 45 individuals admitted to rehabilitation programs. The correlation of .93 was reported between SIT and WAIS-FS IQ scores.

The preliminary evidence indicates that the SIT yields valid IQ scores for children and adults and furnishes a useful screening instrument for an evaluation of mental ability. The SIT was given in order to determine its relationship to the CB-SIT which was derived from it. Both the CB-SIT and the SIT were compared with the WAIS.

Wechsler Adult Intelligence Scale

The WAIS (Wechsler, 1955) is a revision of the original Wechsler-Bellevue Intelligence Scale. It was designed to be a broad-sampling, wide range of ability, individual test of adult intelligence. Norms are provided for ages 16 to 75 and for IQ's ranging from 45 to 159 for the young adult.

In contrast to the Terman-Binet plan of grouping items according to difficulty, mixing content randomly, Wechsler arranges then into subtests of various types. There are 11 scored subtests, grouped in Verbal and Performance scales. The Verbal scale includes tests of Information, Comprehension, Digit Span, Similarities, Arithmetic, and Vocabulary. The Performance scale includes Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol tests. These scales yield three scores-verbal, performance, and a combination - which becomes the full-scale IQ.

The WAIS has been repeatedly correlated with the Stanford-Binet Scale (Terman-Merrill, 1960) as well as other well-known tests of intelligence. Several

summaries of such research have been published (Guertin, Frank & Rabin, 1956; Guertin, Rabin, Frank, & Ladd, 1962; Guertin, Ladd, Frank, Rabin, & Hiestes, 1966). Correlations with the Stanford-Binet in unselected adolescent or adult groups generally fall into the .80's and .90's. Within more homogeneous samples, such as college students, the correlations tend to be considerably lower. Group tests yield somewhat lower correlations with the WAIS, such correlations ranging from about .40 to about .80.

The WAIS was chosen as the criterion measure to evaluate the CB-SIT and the SIT. For adults, the WAIS is more extensively used than the Stanford-Binet Scale (DeCecco, 1967). In addition, most of the validity studies of the SIT have used elementary level, mentally retarded Ss. Availability of well-established norms for adults of different ages made the WAIS a logical criterion measure.

State-Trait Anxiety Inventory

The State-Trait Anxiety Inventory (STAI) developed by Spielberger, et al. (1970) consists of separate self-report scales for measuring two distinctive types of anxiety: state anxiety (A-State) and trait anxiety (A-Trait). A-State is conceptualized as a transitory emotional state or condition of the S that varies in intensity and fluctuates over time. It has been demonstrated that scores on the STAI A-State scale increase in response to various kinds of stress and decrease as a result of relaxation training (Spielberger, et al., 1970). Further evidence bearing on the construct validity of the STAI A-State scale may be found in recent studies by O'Neil (1970) and O'Neil, Spielberger, and Hansen (1969).

The five-item A-State scale of the STAI was used to measure levels of state anxiety before and after all intelligence testing sessions. This scale was chosen in order to investigate the relationship between levels of state anxiety and the test administration method. This short form A-State scale consists of the five items from the 20-item A-State scale having the highest item-remainder correlations for the normative sample of the STAI. In a series of studies using this scale, O'Neil (1970), Lawler, O'Neil, and Hansen (1970), and Leherissey, O'Neil, and Hansen (1970) report ranges of alpha reliabilities from .81 to .89. A copy of this scale is presented in Appendix A.

A-Trait refers to a relatively stable individual difference in anxiety proneness, i.e., to differences in the disposition or tendency to respond with elevations in A-State in situations that are perceived as threatening. Spielberger, et al. (1970) report relatively high correlation coefficients of validity with the Taylor Manifest Anxiety Scale (TMAS) and the Institute for Personality and Ability Testing Anxiety Scale (IPAT).

The State-Trait Anxiety Inventory A-Trait Scale (Appendix B) was administered at the beginning of the testing procedures because the levels of observed A-State scores will be partly a function of the anxiety proneness of the student.

Attitude Inventory

An adjective scale modeled after the Semantic Differential (Osgood, et al., 1957) was used to measure attitudes toward the concepts of "computer testing" and "examiner testing." This scale was given in the future tense form for the pretest of each testing session and the past tense form for each posttest session (Appendix C). Mathis, Smith, and Hansen (1969) have reported test-retest reliabilities ranging from

.88 to .95. More importantly, these scale scores have been shown to be related to performance in a computer-assisted instruction mode (Mathis, et al., 1969). Students who committed many errors on the CAI program rated the computer as more dull, bad, worthless, weak, unpleasant, boring, and depressing than did students who made fewer errors.

This measure was selected in order to study the relationship between student test attitudes and the test administration method.

Examiner Rating Scale

An adjective scale was developed from examiner ratings of test behavior from the Stanford-Binet Scale (Terman-Merrill, 1960). This measure (Appendix D) was completed by the WAIS and SIT examiners at the conclusion of each test administration. The purpose of this instrument was to explore the relationship between examiner test behavior ratings and student self-report attitude and anxiety measures.

Subjects

Forty-eight undergraduate students were obtained from the introductory psychology and educational psychology courses at Florida State University. Ss were randomly assigned to the three experimental testing conditions with an equal number of males and females within each group. Participation in the investigation served to fulfill certain course requirements for the students.

Apparatus

The computer-based SIT was administered via an IBM 1500 Instructional System (IBM, 1967). Terminals for this system consist of a cathode ray tube (CRT), a light pen, and a keyboard. These terminals are located in an air-conditioned, sound-deadened room. The CAI system administered the STAI A-State anxiety scales both before and after the CB-SIT. Student input for the CB-SIT consisted solely of typewritten responses. All student responses were automatically recorded and evaluated by the computer system.

Standard testing materials were employed for the WAIS and SIT test administrations.

Experimental Design

A Latin square design (Winer, 1962) was employed to counterbalance potential order and sequence effects of the three different intelligence test administrations. Each S was individually tested with the WAIS, SIT, and CB-SIT. A time interval of approximately one week separated the different test administrations. In addition, each testing session included an administration of the attitude scale and STAI A-State anxiety scale both before and after each testing period. Figure 1 exemplifies the design of the present investigation.

Ss were randomly assigned to the three experimental testing groups with an equal number of males and females within each group. In order to control for possible examiner bias effects, there was a random assignment of the WAIS and SIT examiners to students.

Group	Week 1	Week 2	Week 3
Group 1 (n=16)	WAIS	CB-SIT	SIT
Group 2 (n=16)	CB-SIT	SIT	WAIS
Group 3 (n=16)	SIT	WAIS	CB-SIT

Fig. 1—Experimental design with number of Ss

Experimental Procedures

All three experimental testing sessions involved the following sequence of events: (1) a pretask period during which Ss responded to the attitude and state anxiety scales; (2) a task period in which one of the three intelligence tests was administered; and (3) a posttask period in which Ss responded to the attitude and state anxiety scales. These procedures are described in detail below.

Pretask Period

All Ss reported to the Computer-Assisted Instruction Center for all three experimental testing sessions. Initially, the STAI A-Trait Inventory was completed by all Ss before any other testing was done.

Prior to receiving terminal instructions for the CB-SIT, Ss were asked to respond to the attitude scale concerning computer testing. The instructions were given in the future tense and and asked the students to react to "Computer testing will be ...?" Ss were then asked to read a description of the operation of the CAI terminal. They were then given practice in the operation of the terminal keyboard and instructed how to execute erase and enter functions.

After "signing on," Ss immediately responded to the STAI A-State scale. The instructions for the A-State scale given before each testing period asked the Ss to "tell how you feel right now." Following this, a number of practice frames were presented to familiarize the Ss with the CB-SIT program characteristics and operations.

Similar procedures were followed for the examiner testing sessions (WAIS and SIT). After the examiner lead the Ss to the testing room, the S responded to the attitude and anxiety scales which were presented in conventional paper and pencil fashion before the actual intelligence testing began.

Task Period

During the task period Ss were individually examined with one of the three intelligence tests. The order of the test administrations was counter-balanced for the three experimental groups. Figure 1 presents a description of the experimental design of the investigation.

The CB-SIT was administered via an IBM 1500 Instructional System. Ss read the items presented on the CRT and responded by typing in their response for immediate

computer evaluation. As part of experimental procedures, all Ss responded to the entire set of test items from the ages of 12 to 27.

The administrations of the WAIS and SIT conformed carefully to the standard procedures as described in their respective test manuals. The testing rooms were free from distracting noises and intrusions. These rooms are well lighted and ventilated. Sufficient time was scheduled for the testing so that good rapport could be established and maintained and that the administration could proceed in an easy, unhurried manner.

The WAIS examiners were five graduate students from the Department of Psychology at the Florida State University. The four SIT examiners were graduate students from the Department of Educational Research and Testing at Florida State University. All examiners had extensive experience with their particular test instruments.

Posttask Period

After each **S** completed each of the three intelligence tests, he was given the anxiety and attitude scales. The computer presented the STAI A-State scale upon completion of the CB-SIT, whereas, the attitude scale was presented in paper and pencil fashion. For the examiner administered tests (WAIS and SIT), both the attitude and anxiety scales were presented in paper and pencil format.

The attitude scale was presented in past tense form, i.e., "Computer testing was ..." or "Examiner testing was ..." Retrospective instructions were used for the STAI A-State scale, i.e., "tell me how you felt during the test which you just completed." Following each examiner-administered test, the examiner completed the examiner rating scale for the **S** just tested.

Upon completion of the experimental procedures for all tests, the Ss were debriefed, given some additional information concerning the general nature of the experiment, and cautioned not to discuss the experiment with other Ss who have not yet completed the procedures. Finally, an extended tour of the CAI Center was offered to the students.

CHAPTER IV
RESULTS

The results section is organized into eight major sections. In the first section, IQ performance scores were evaluated as a function of test administration procedures. Second, the internal consistency reliabilities of the CB-SIT and SIT were evaluated. The third section deals with concurrent validity analyses for the CB-SIT and SIT with the WAIS. Predictive validity analyses of the WAIS, SIT, and CB-SIT in relation to college grade point average are presented in the fourth section. Evaluations of STAI A-State scores and attitude scale scores are presented in the fifth and sixth sections. The seventh section presents the analyses of the examiner rating scale scores in relation to performance, state anxiety, and attitudes. Finally, methodological problems and suggestions for future revisions in the presentation and scoring strategies for the CB-SIT are presented.

Evaluation of Test Administration Procedures

The means and standard deviations for the WAIS Full Scale, SIT, and CB-SIT computer-dervied IQ scores for this sample of college students (N=48) are presented in Table 1. As may be seen in Table 1, the mean WAIS FS-IQ scores were fairly consistent across all three experimental groups. However, CB-SIT and SIT IQ scores did not show the same consistent relationship. In Groups I and II there was a consistent increase in SIT IQ scores when the CB-SIT preceded the SIT, whereas in Group III,

TABLE 1
Means and Standard Deviations of WAIS, SIT, and CB-SIT IQ Scores by Entire Sample and for Experimental Groups

TEST	WAIS VIQ	WAIS PIQ	WAIS FSIQ	CB-SIT	SIT
Group I (n=16)					
Mean	121.4	114.5	119.7	123.7	135.4
SD	5.7	9.6	5.9	8.6	9.4
Group II (n=16)					
Mean	123.6	115.5	121.3	119.5	132.4
SD	7.0	7.4	6.0	8.4	9.3
Group III (n=16)					
Mean	122.8	115.4	121.0	120.2	127.1
SD	8.6	11.4	8.1	8.7	9.1
Entire Sample (N=48)					
Mean	122.6	115.1	120.7	121.2	131.6
SD	7.1	9.4	6.7	8.6	9.7

SIT IQ scores were higher than CB-SIT IQ scores when the SIT preceded the CB-SIT. A 3 x 3 latin square repeated measures analysis of variance was employed to evaluate these effects of tests and testing sessions (Winer, 1962). The dependent variable was the obtained IQ scores from the WAIS, SIT, and CB-SIT.

The results of this analysis (Table 2) revealed a statistically significant difference between the mean IQ scores derived from the WAIS, SIT, and CB-SIT ($F=74.9$, 2 and 90 df, $p<.01$). The SIT yielded the highest IQ score of all three tests ($\bar{X}$ = 131.6). The mean WAIS FS-IQ ($\bar{X}$ = 120.7) and CB-SIT IQ ($\bar{X}$ = 121.2) scores were consistent with each other. In addition, a significant day of test administration effect was found (F = 8.38, 2 and 90 df, p <.01). The mean IQ scores for the three days of testing were 122.1, 125.7, and 125.6, respectively. Test administration order and the interaction of tests x administration days were not found to be statistically significant.

TABLE 2
Analysis of Variance for Tests and
Test Administration Procedures

Source	df	MS	F
Between Subjects			
Groups	2	147.84	.95
Subj. w/groups		155.94	
Within Subjects			
Test Sessions (A)	2	205.06	8.38**
Tests (B)	2	1831.18	74.87**
A x B	2	24.26	.99
Error (within)	90	24.46	

**p <.01

Reliability Analysis of the CB-SIT and SIT

To determine the reliability of the CB-SIT computer-derived IQ scores and the SIT IQ scores, a modified split-half procedure was employed in the following manner. The basal items (ten in a row correct) and the ceiling items (ten in a row incorrect) were ignored in the odd-even item determination. The first item to be counted was the first error item following the basal age item. From that point, the odd numbered items were counted separately until the first of the ten ceiling items was encountered. If a ceiling was not observed for an S, all items following the basal age were counted including the final item of the test.

Beginning with the second item following the basal age item, a similar procedure was employed to determine the even-numbered score. When an odd number

of items occurring within the effective scoring range was encountered, the final item was deleted from the analysis in all cases. Pearson product-moment coefficients were used to determine the relationship between the sum of the correct answers for the odd-even divisions of the CB-SIT and SIT.

Table 3 presents the modified split-half coefficients for the CB-SIT and SIT for the total sample and for each of the three experimental testing groups. In all cases, the SIT yielded higher reliabilities than the CB-SIT computer-derived IQ scores.

TABLE 3
Modified Split-Half Reliability Coefficients for the
CB-SIT and SIT by Entire Sample
and for Experimental Groups

Group	CB-SIT	CB-SIT*	SIT	SIT*
Total Sample (N=48)	.66	.79	.79	.88
Group I (n=16)	.72	.83	.78	.87
Group II (n=16)	.46	.64	.65	.79
Group III (n=16)	.76	.86	.87	.95

*NOTE: These correlations have been adjusted for test length.

The Spearman-Brown correction for test length procedure was applied to the half test estimates. As can be noted, the range of adjusted coefficients for the SIT was .79 to .95. The range of the adjusted CB-SIT reliability estimates was from .64 to .86. Due to the variable number of items across subjects, the Spearman-Brown adjustment procedure may have resulted in an underestimation of the total reliability of the test.

Concurrent Validity Evaluation

Pearson product-moment coefficients were used to determine the concurrent validity relationships between the WAIS, SIT, and CB-SIT (computer-derived scores). Since a significant effect of day of administration was found, the concurrent validity data will be presented for both the entire sample, as well as for the three experimental testing orders, i.e., Groups I, II, and III. In addition, relationships of the CB-SIT and SIT to WAIS subtest scores were investigated.

Table 4 presents the correlation coefficients between the WAIS, SIT, and CB-SIT for the entire sample of college students. These results indicate that the CB-SIT and SIT IQ scores correlated positively with each other ($r=.75$, $p < .01$). In addition, both the CB-SIT and SIT correlated positively with the WAIS VIQ's ($p < .05$), PIQ's ($p < .05$), and FS-IQ's ($p < .01$)

TABLE 4
Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Entire Sample (N=48)

TEST	CB-SIT	SIT	WAIS VIQ	WAIS PIQ	WAIS FS-IQ
CB-SIT		.75**	.54**	.32*	.54**
SIT			.48**	.32*	.52**
WAIS VIQ				.20	.77**
WAIS PIQ					.78**
WAIS FSIQ					

*p < .05 **p < .01

The magnitude of the differences between the correlations of the CB-SIT and SIT with the WAIS were tested for statistical significance using a series of Fisher t tests for correlated samples (McNemar, 1962). The comparison of the correlation between the CB-SIT and WAIS FS-IQ's with the correlation of the SIT and WAIS FS-IQ's resulted in a non-significant t ratio of .23, df = 45. In addition, similar analysis showed no significant differences in the correlations between CB-SIT and SIT scores with WAIS VIQ's or PIQ's. Thus, the magnitudes of the validity relationships were essentially parallel for the CB-SIT and SIT in relation to the WAIS IQ scores.

Since the standard deviations for the entire sample (see Table 1) reflected a restriction in range, estimated coefficients were derived for the entire range of subjects (Guilford, 1956). The resulting correlations of .92 between CB-SIT and WAIS FS-IQ's and .89 between SIT and WAIS FS-IQ's demonstrated that the correlations of .54 and .52 underestimated the concurrent validity between these intelligence tests. However, the correlations of .54 and .52 probably adequately reflect the nature and magnitude of the relationship for a college population.

The consistency of the correlational relationships of the CB-SIT and SIT IQ scores with the WAIS IQ scores can also be seen by comparing these two SIT scores with the 11 WAIS subtest scale scores. Table 5 presents these correlational coefficients.

For the WAIS verbal subtests, CB-SIT computer-derived IQ scores correlated significantly with Information, Arithmetic, Similarities, and Vocabulary scale scores. With the Performance subtests of the WAIS, the CB-SIT correlated positively with the Block Design subtest scale scores. In comparison, the SIT correlated significantly with the verbal subtests scores for Information, Similarities, and Vocabulary. Within the

TABLE 5

Coefficient of Correlation Between CB-SIT, SIT and WAIS Subtests for Entire Sample (N=48)

	WAIS Verbal Subtests						WAIS Performance Subtests				
	Information	Comprehension	Arithmetic	Similarities	Digit Span	Vocabulary	Digit Symbol	Picture Completion	Block Design	Picture Arrangement	Object Assembly
CB-SIT	.43**	.06	.40**	.30**	.11	.53**	.03	.19	.39**	.22	.19
SIT	.45**	.20	.15	.29**	.09	.35*	.08	.18	.42**	.30*	.16

*p < .05
**p < .01

Performance Section, the SIT correlated with Block Design and Picture Arrangement subtest scale scores.

The magnitude of these correlational differences between the CB-SIT and SIT IQ scores and the WAIS IQ scores were evaluted with a series of Fisher t tests for correlated samples. The results of these analyses revealed non-significant correlational differences between the CB-SIT and SIT IQ scores for nine of the 11 WAIS subtest scale scores. However, the CB-SIT was shown to correlate more positively with the Vocabulary subtest of the WAIS than the SIT ($t=2.67$, 45 df, $p< .05$). In addition, the CB-SIT IQ scores were found to relate significantly higher than SIT IQ scores with the arithmetic subtest scale scores of the WAIS ($t=2.02$, 45 df, $p < .05$).

Due to the significant day of test administration effect, separate correlational analyses were computed for each of the three testing groups. These correlational coefficients are presented in Tables 6, 7, and 8. In Groups I and III, CB-SIT and SIT scores correlated positively with WAIS FS-IQ's. However, these correlations were not significant for Group 2. The patterns of validity relationships seem to be consistent with the CB-SIT correlating about equally as the SIT with WAIS VIQ's, PIQ's, and FS-IQ's. In addition, the results of a series of Fisher t tests indicated that the correlations between the CB-SIT and SIT with WAIS, VIQ's, PIQ's, or FS-IQ's were not significantly different from each other within Groups I, II, and III.

TABLE 6
Coefficients of Correlation Between WAIS SIT, and CB-SIT for Group I (n=16)

TEST	CB-SIT	SIT	WAIS VIQ	WAIS PIQ	WAIS FS-IQ
CB-SIT		.77**	.42	.39	.51*
SIT			.49	.42	.60*
WAIS VIQ				.12	.68**
WAIS PIQ					.80**
WAIS FS-IQ					

* $p< .05$
** $p < .01$

TABLE 7
Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Group II (n=16)

TEST	CB-SIT	SIT	WAIS VIQ	WAIS PIQ	WAIS FS-IQ
CB-SIT		.79**	.54*	-.05	.36
SIT			.45	.07	.37
WAIS VIQ				.17	.80**
WAIS PIQ					.72**
WAIS FS-IQ					

* $p < .05$
** $p < .01$

TABLE 8
Coefficients of Correlation Between WAIS, SIT, and CB-SIT for Group III (n=16)

TEST	CB-SIT	SIT	WAIS VIQ	WAIS PIQ	WAIS FS-IQ
CB-SIT		.75**	.75**	.55*	.82**
SIT			.69**	.52*	.76**
WAIS VIQ				.25	.78**
WAIS PIQ					.80**
WAIS FS-IQ					

* $p < .05$
** $p < .01$

Predictive Validity of the WAIS, SIT, and CB-SIT in Relation to College Grade Point Average

Multiple regression analyses were conducted in order to determine the predictive validity of the three intelligence tests in relation to College Grade Point Average (GPA). The GPA averages were obtained from the Office of the Registrar at Florida State University. Since sex of the subject correlated significantly with the criterion of GPA ($r=.51$, $p < .01$), the analyses were computed with sex as a predictor. The following four linear regression models were compared in terms of their predictive efficiency:

Model I	GPA	$=B_0 + B_1$ Sex + Error
Model II	GPA	$=B_0 + B_1$ Sex + B_2 CB-SIT + Error
Model III	GPA	$=B_0 + B_1$ Sex + B_3 SIT + Error
Model IV	GPA	$=B_0 + B_1$ Sex + B_4 WAIS + Error

Table 9 shows these regression equations with the dependent variable (GPA), the multiple correlation coefficient (R), and the variance estimates (R^2).

TABLE 9
Multiple Regression Analysis with GPA As the Criterion (N=48)

	Equation	R	R^R2
I	GPA = 3.13 - .62 (Sex)	.55**	.26
II	GPA = -.57 - .66 (Sex) + .03 (CB-SIT)	.66**	.44
III	GPA = .16 - .70 (Sex) + .02 (SIT)	.62**	.38
IV	GPA = .40 - .65 (Sex) + .02 (WAIS FS-IQ)	.56**	.32

** $p < .01$

It should be noted that in all of the equations, sex was negatively related to the dependent variable. In other words, females tended to show higher GPA's than the males in this particular sample of college students. In addition, the multiple R's were significant ($p < .01$) for all of the four regression models.

To determine the significance of the variance contribution of each of the intelligence tests over and above that of the sex factor (26 percent), F tests were calculated which compared Equation 1 with Equations 11, 111, and 1V. This procedure

enables one to estimate the contribution of each of the IQ tests independently of the effect of sex.

Equation 11 with sex and CB-SIT IQ scores accounted for 44 percent of the variance and yielded a significant F of 14.1, 1 and 45, df ($p < .01$) when compared to Equation 1. Thirty-eight percent of the variance was accounted for by Equation 111 which included sex and SIT IQ scores as predictors. When this equation was compared to Equation 1, a significant F of 9.12, 1 and 45, df ($p < .01$) was shown. However, Equation IV with sex and WAIS FS-IQ scores as predictors did not differ significantly from Equation 1 (F= 3.88, 1 and 45, df).

Further evidence concerning the effectiveness of the prediction equations can be seen by comparing their respective indices of forecasting efficiency (Guilford, 1956). The index of forecasting efficiency is defined as the percentage reduction in errors of prediction by reason of correlation between variables. The forecasting efficiency (E) for a given equation is given by the formula, $E = 100(1 - \sqrt{1 - R^2})$.For Equation 1, E was equal to 14 percent. This indicated that GPA predictions based on this equation would be 14 percent better than those made merely from a knowledge of the mean GPA for this sample. Equation 11 with sex and CB-SIT IQ scores as predictors yielded a 25.2 percent efficiency index. Equation 111 with sex and SIT IQ scores showed an E of 21.3 percent. Finally, an E index of 17.5 percent was found for the equation which included sex and WAIS FS-IQ scores.

Evaluation of State Anxiety

In order to investigate the affective nature of the computer and examiner testing procedures, the STAI A-State measures were examined as a function of test session, intelligence test, and pre-post measurement periods. This statistical analysis of the A-State data was based on the short form of the STAI A-State scale. The same five items were given both before and after the administration of each intelligence test.

Table 10 presents Kuder-Richardson 20 reliability estimates for the STAI A-State scale scores for the different measurement periods. As can be noted, the range of the coefficients was from .86 to .92 for this sample of college students (N=48). These results are consistent with the normative data reported on the total scale (Spielberger, et al., 1971).

The means and standard deviations of the STAI A-State scores by intelligence tests are presented in Table 11. These data were evaluated by a 3 x 3 x 2 latin square analysis of variance with repeated measures on the last 2 factors. The independent variables were intelligence tests (CB-SIT, SIT, WAIS), test session (first week, second week, third week), and testing periods (pre-post). STAI A-State scores were the dependent variable. The results of this analysis are presented in Table 12.

Results of this analysis revealed that state anxiety scores were significantly higher in the computer testing situation ($\bar{X} = 11.4$) in comparison to the two examiner testing conditions (WAIS $\bar{X} = 9.4$; or SIT $\bar{X} = 9.4$) as indicated by the significant test effect. In addition, state anxiety scores decreased significantly across the three testing sessions ($\bar{X}_1 = 10.8$; $\bar{X}_2 = 10.1$; $\bar{X}_3 = 9.3$). Moreover, a small increase in state anxiety scores was observed from the pre ($\bar{X} = 9.8$) to post ($\bar{X} = 10.3$)

TABLE 10
Kuder-Richardson 20 Reliability Estimates
for the STAI A-State Measures

TEST	Pre	Post
WAIS	.92	.87
SIT	.86	.89
CB-SIT	.91	.92

TABLE 11
Means and Standard Deviations of STAI A-State
Scores by Intelligence Test for Experimental
Groups and Entire Sample

TEST	CB-SIT		SIT		WAIS	
	pre	post	pre	post	pre	post
Group I (n=16)						
Mean	10.81	13.25	9.63	9.13	8.94	11.69
SD	4.02	3.99	3.16	2.73	3.15	3.30
Group II (n=16)						
Mean	11.75	12.75	9.50	8.44	8.88	8.25
SD	3.26	3.53	3.03	3.65	3.58	3.21
Group III (n=16)						
Mean	9.75	10.19	9.44	10.00	9.31	9.31
SD	3.34	4.61	3.39	3.81	4.01	2.80
Entire Sample (N=48)						
Mean	10.77	12.06	9.52	9.19	9.04	9.75
SD	3.57	4.20	3.13	3.42	3.52	3.37

TABLE 12
Analysis of Variance of STAI A-State Scores
by Intelligence Tests, Test Sessions, and
Pre-Post Measurement Periods

Source	df	MS	F
Between Subjects			
Groups	2	20.90	.52
Subj. w/groups	45	40.27	
Within Subjects			
Intelligence Tests (A)	2	133.43	19.91**
Test Sessions (B)	2	51.21	7.64**
(A x B)'	2	7.76	1.16
Pre-Post Periods (C)	1	22.22	3.32
C x Groups	2	20.15	3.01
A x C	2	16.26	2.43
B x C	2	16.84	2.51
(A x B)' x C	2	2.76	.41
Error (within)	225	6.70	

** $p < .01$

measurement periods; however, this difference only approached significance ($p < .10$). No second or third order interactions were found to be statistically significant.

A subsidiary post hoc analysis which disregarded testing order was computed for the STAI A-State scores as a function of sex, level of trait anxiety, and intelligence tests. Ss were divided into the categories of High Trait (HA) and Low Trait (LA) anxious based upon a median split on the A-Trait scores within sex. The medians for males and females were 35.5 and 35.5, respectively. In addition, the mean A-Trait score for HA males was 44.3, whereas the mean A-Trait score for HA females was 42.3. Mean A-Trait scores for LA males and females were 29.9 and 28.8. Four Ss were deleted from the analysis in order to insure comparability of the A-Trait means and medians. The means and standard deviations of the STAI A-State scores as a function of trait anxiety, sex, and intelligence tests are presented in Table 13.

Results of this 2 x 2 x 3 x 2 analysis of variance with repeated measures on the last two factors revealed a significant main effect for the different intelligence tests ($F=11.1$, 2 and 80 df, $p < .01$). In accordance with the previous analysis, the CB-SIT yielded the highest anxiety levels in comparison to the SIT and WAIS. Sex of the S and level of A-Trait were not found to be differentially related to the observed A-State scores. In addition, A-State scores were not found to increase from the pre to post measurement periods. None of the interactions were found to be statistically significant.

TABLE 13
Means and Standard Deviations of STAI A-State
Scores as a Function of Trait Anxiety
Sex, and Intelligence Tests

GROUPS	CB-SIT		SIT		WAIS	
	Pre	Post	Pre	Post	Pre	Post
HA Males						
Mean	11.27	12.55	9.82	9.82	10.55	10.91
SD	5.18	5.94	3.71	4.05	3.80	3.45
HA Females						
Mean	11.27	12.00	9.82	8.64	8.82	10.36
SD	3.07	3.77	2.75	3.32	3.09	3.64
LA Males						
Mean	9.45	10.91	9.27	9.64	7.82	8.91
SD	3.01	4.50	3.04	2.46	2.48	2.98
LA Females						
Mean	10.82	12.36	9.27	8.73	8.64	8.45
SD	3.03	1.96	3.52	4.34	4.70	3.36

TABLE 14
Kuder-Richardson 20 Reliability Estimates
For the Attitude Scale Measures

TEST	Pre	Post
WAIS	.80	.93
SIT	.81	.83
CB-SIT	.81	.90

Evaluation of Student Attitudes

To further evaluate the affective nature of the computer and examiner testing procedures, the attitude scale scores were analyzed as a function of test session, intelligence test, and pre-post measurement periods. The statistical analysis of the attitudinal data was based on the total score for the 15-item attitude scale. The same 15 items were given both before and after the administration of each intelligence test.

Table 14 presents Kuder-Richardson 20 reliability estimates for the attitude scale scores for the two measurement periods. As can be noted, the range of the coefficients was from .80 to .93 for this sample of college students (N=48). These data favorably compare with the results reported by Mathis, et al. (1969).

The means and standard deviations of the attitude scale scores are reported in Table 15 for the three experimental groups, for the three intelligence tests, and for the pre-post measurement periods. These data were evaluated by a 3 x 3 x 2 repeated measures analysis of variance. The results of this analysis are presented in Table 16 in which the most important finding was the significant triple interaction. This text x test session x pre-post period interaction indicated that student attitudes were differentially influenced by the experimental conditions. Student attitudes toward WAIS examiner testing increased from pre to post evaluations regardless of the testing session, whereas student attitudes toward the computer testing experience decreased in all cases. Student attitudes toward the SIT examiner testing increased when the SIT was administered on the second or third testing session, but decreased when the SIT was the first test encountered by the S. In addition, the test x pre-post period interaction was significant.

TABLE 15
Means and Standard Deviations of Attitude Scale Scores
by Intelligence Test for Experimental
Groups and Entire Sample

TEST	CB-SIT		SIT		WAIS	
	Pre	Post	Pre	Post	Pre	Post
Group I (n=16)						
Mean	73.50	64.19	68.81	70.25	70.75	76.88
SD	10.93	11.83	10.09	10.43	11.03	11.06
Group II (n=16)						
Mean	73.56	67.13	68.19	73.69	71.75	76.31
SD	6.97	11.60	7.11	8.17	7.60	6.57
Group III (n=16)						
Mean	72.56	69.38	73.06	70.50	70.25	73.38
SD	9.86	17.51	10.85	12.93	11.14	12.97
Entire Sample (N=48)						
Mean	73.21	66.90	70.02	71.48	70.92	75.52
SD	9.21	13.79	9.54	10.57	9.86	10.43

The results of the ANOVA further indicated that student attitudes were more positive toward the WAIS ($\bar{X}$ = 72.9) than toward either the CB-SIT ($\bar{X}$ = 70.1) or the SIT ($\bar{X}$ = 70.7) as indicated by the significant main effect for intelligence tests. The main effects of testing sessions or pre-post periods were not statistically significant.

TABLE 16
Analysis of Variance of Attitude Scale Scores by Intelligence Tests, Test Sessions, and Pre-Post Measurement Periods

Source	df	MS	F
Between Subjects			
Groups	2	26.10	.061
Subj. w/groups	45	428.29	
Within Subjects			
Intelligence Tests (A)	2	212.52	4.04*
Test Sessions (B)	2	79.80	1.52
(A x B)'	2	73.02	1.39
Pre-Post Periods (C)	1	1.13	.21
C x Groups	2	21.79	.42
A x C	2	831.26	15.81**
B x C	2	22.89	.44
(AB') x C	2	165.66	3.15*
Error (within)	225	52.57	

* $p < .05$
** $p < .01$

Examiner Rating Scale Analysis

Total scores from the 12-item, 7 point Likert-type adjective scale developed from examiner ratings of test behavior from the Stanford-Binet Scale, Form L-M (Terman-Merrill, 1960) were analyzed for correlational relationships with the student self-report attitude measures, anxiety measures, and observed intelligence test performance measures. For the WAIS, the mean examiner rating scale score was 53.6 with a standard deviation of 9.8. Table 17 presents the correlational relationships with these scores.

As will be noted from Table 17, female students were rated higher than the male students by the WAIS examiners as indicated by the significant negative correlation

with sex of the student. Moreover, performance during the testing session was related to examiner reports as indicated by the significant correlations with WAIS VIQ's and FS-IQ's. WAIS PIQ scores were not found to be related to the total examiner rating scores. In terms of the affective measures, the WAIS examiner ratings were not related to either the pre WAIS attitude or anxiety scale scores. However, significant relationships were found for the post attitude and anxiety measures. The examiner ratings correlated positively with the student total attitude scale scores. Students who positively rated the testing experience were also rated more favorably by their particular examiners. In addition, the ratings were negatively related to the STAI post A-State scores. Higher ratings were given for students who reported lower A-State scores.

TABLE 17
Coefficients of Correlation for Total
WAIS Examiner Rating Scale Scores

	r
Sex	-.32*
WAIS VIQ	.33*
WAIS PIQ	.27*
WAIS FS-IQ	.39*
Pre WAIS Attitude Scale	.16
Pre WAIS Anxiety Scale	-.07
Post WAIS Attitude Scale	.30*
Post WAIS Anxiety Scale	-.34*

* $p < .05$

TABLE 18
Coefficients of Correlation for Total
SIT Examiner Rating Scale Scores

	r
Sex	-.27
SIT IQ	.48**
Pre SIT Attitude Scale	.08
Pre SIT Anxiety Scale	.10
Post SIT Attitude Scale	.20
Post SIT Anxiety Scale	.07

** $p < .01$

Similar findings were observed for the examiner ratings obtained from the SIT testing. The mean SIT examiner rating was 55.6 with a standard deviation of 9.8. Table 18 presents the correlational relationships for these SIT ratings.

Sex of the student was negatively related, but not significantly with the SIT examiner ratings. As with the WAIS, the examiner ratings were positively related to the observed performance of the student. However, the correlations with student attitudes and anxiety scores were not significant for both the pre and post measurement periods.

Contrary to expectation, the correlation for the two sets of examiner ratings was essentially zero. In other words, WAIS examiner rating scores were unrelated to the SIT examiner rating scores for the same students.

Developmental Problems and Suggestions for Future Revisions

Within any programmatic test automation program, revisions are necessitated through the collection of empirical data which reveal certain methodological problems. In the context of the CB-SIT presentation and answer programming strategies, both empirical and anecdotal data from the present study have revealed the necessity of certain modifications in the CB-SIT presentation strategy and in the scoring algorithms.

Test Presentation Strategy Analysis and Modifications

The present CB-SIT program required all students to respond to the entire set of test items from the ages of 12 to 27. The primary rationale for this presentation strategy was the need to collect multiple item response protocols to facilitate future item dictionary revisions. However, data from the present investigation revealed a number of adverse consequences of such a procedure. First, this procedure substantially lengthened the amount of time required to complete the CB-SIT (50 minutes). Students responded to a number of items that did not enter into the actual calculation of the IQ score ($\bar{X}$ = 18.2 items). For a bright student, this number could be substantial. In addition, students were required to respond to all item parts within the item. For example, certain test items contain four or five subquestions. If a certain number of the item parts were answered incorrectly, it would be unnecessary to present the remaining component questions. Figure 2 presents a description of such an item. However, in this investigation, the students were required to respond to all of the question components.

The amount of system time allowed for item responses was programmed to be 2 minutes for the verbal items and 1 minute for the mathematical questions. If time limits were exceeded for an item, i.e., "timed out," the students were instructed to "please answer the question or type pass." They were then given another opportunity to answer the question. If two time outs occurred at a particular item, the item was scored as incorrect and the program continued to the next item. In addition, if the student was in the middle of typing his response, the response would be erased when he was given the comment to proceed faster. Then the student would have to retype his entire response.

What parts of an animal's body are indicated by the following?

A. Caudal
B. Dorsal
C. Ventral
D. Cephalic
E. Oral
F. Mouth

Figure 2.--Item 24-9 from the SIT.

Analysis of item "time outs" revealed that these system time limits may not have been appropriate in all instances. Of the 17 items with a high number of time outs, 13 items were math in nature and only 4 required word, phrase or sentence input. This would seem to indicate that the amount of possible retyping of an answer would be at a minimum due to the larger proportion of math items which only require a numerical answer. However, student comments revealed this procedure to be somewhat frustrating in nature.

The nature of the student input, i.e., word, phrases, etc. dictated the inclusion of a typing skill factor. Some students indicated that they were frustrated to a certain extent because of their difficulty in typing their answers. Within this context, certain students also indicated their inability to change answers with ease. In other words, unfamiliarity with the typewriter, plus programming methodology, seemed to cause certain adverse affective reactions. Other adverse reactions seemed to be precipitated by the lack of supportive feedback within the computer testing experience.

Another area of difficulty centered around the particular questionning procedure for partially correct answers. Students were instructed to "Tell me more" by the computer when certain answers were scored as partially correct. As with the time out procedure, when a student was instructed to expand his answer, the student would then have to retype another entire answer. Anecdotal student comments revealed that the students perceived this "Tell me more" message to induce nervousness, frustration, and to be too mechanical in nature.

Based upon the above considerations, major program revisions are being implemented to amend the observed problem areas. First, the presentation strategy is being revised to incorporate the concepts of basal age and ceiling age in the administration of the CB-SIT. The program for college subjects will begin at age test item 21-0 and proceed sequentially in a reverse order until the basal age or "ten in a row correct" is observed. The program will then branch back to item 21-3 and proceed until the ceiling (ten in a row incorrect) is observed. The program will terminate when the ceiling or final test item is observed. This revised presentation procedure should substantially reduce the number of items to which the S will have to respond.

In addition, intra-item presentation changes are being made to reduce the total number of test responses. Within multiple response items (items requiring 2, 3, 4, 5, or 6 individual answers), items will be presented as long as the scoring cut-off for

an incorrect answer to the total item has not been reached. An example will clarify this procedure. Figure 2 presents a description of item 24-9 from the SIT.

In order to obtain a correct answer to this item, the student must correctly identify five out of the six components. Two incorrect responses would indicate an incorrect response to the total item. The initial CB-SIT program presented all six of the item parts. The new version of the CB-SIT presents the items individually until two incorrect answers are given. When two incorrect answers are observed the student is branched to item 25-0. For example, if the student incorrectly answered parts a and b, he would be immediately branched to the next item. As long as there is only recorded incorrect answer within the item set, the following item will be presented. The procedure is similar for the other multiple response items. This revision should substantially reduce the number of responses needed to assess the CB-SIT IQ scores. In addition, the amount of time to take the CB-SIT should be decreased by this procedure.

In terms of the time allowed for responses, the system time limits will be increased from one minute to two minutes for the 13 math items with a substantial number of overtime responses. Time limits for the four verbal items with a large number of overtime responses will be increased from two to three minutes. These changes should more adequately allow enough response time for the items. In addition, the number of "Please answer the question or type pass." system comments should be reduced by these revisions. This procedure should also reduce the amount of retyping on the part of the student which seemed to cause certain negative affective reactions.

Student instructions are also being revised in order to reduce a certain amount of unnecessary typing on part of the student. When asked to define a particular word, many students would respond by "mutiny means to", or "fragrant means" etc. The typing of these words is unnecessary for the student to obtain a correct answer to the item. The revised instructions will present examples that directly show the appropriate manner in which to respond to this type of test questions.

Scoring Analysis and Modifications

A major question concerns the adequacy of this word/phrase dictionary approach for scoring and deriving valid IQ scores for the CB-SIT. Student test item protocols obtained from detailed computer performance records of the 48 Ss were given to three of the SIT examiners used in the present investigation. The 48 test item protocols were randomly assigned to each of the three examiners so that each examiner scored 16 individual protocols. This scoring procedure will be referred to as the examiner-derived method of evaluation. In order to obtain inter-rater reliability, a total of 12 protocols were concurrently scored by each examiner. Table 19 presents the analysis of variance summary for the examiner reliability analysis (Guilford, 1956).

Results of this analysis revealed a significant difference between the IQ scores for these 12 students, which was expected. The more important finding was the indication of high interexaminer reliability as shown by the nonsignificant examiner scoring effect. Thus, the interrater reliability based upon the above analysis of variance summary was shown to be .73 (Guilford, 1956).

TABLE 19
Analysis of Variance for Examiners

Source	df	MS	F
Students	11	110.71	9.33**
Examiners	2	24.09	2.03
Error	22	11.86	

** $p < .01$

Analysis of examiner-derived versus computer-derived IQ scores yielded a correlation of .89 for the entire sample of 48 college students. This is to be contrasted with the initial field trial results which showed a correlation of .83 between the computer-derived and examiner-derived IQ scores. In the present study the mean computer-derived IQ score was 121.2 (sd=8.56); in comparison, the mean examiner-derived IQ score was 125.7 (sd=9.6). In terms of the number of correct answers for all of the CB-SIT items, the mean number of computer-derived correct answers on the CB-SIT was 36.0, whereas the mean number of correct responses was 39.5 for the examiner-derived answers. The 11 point IQ score point difference observed in the initial field trial between the computer scoring procedures and the criterion was reduced to 4.5 IQ points in the present study.

Thus, the computer scoring procedures for the CB-SIT still tend to produce lower estimates of IQ scores than the criterion. Given this difference in scoring effectiveness, the concurrent and predictive validity analyses were reanalyzed using the examiner-derived CB-SIT IQ scores. Equivalent results were found for the evaluation of test administration procedures, for the concurrent validity evaluation of the CB-SIT, SIT, and WAIS, and for the predictive validity evaluation in terms of GPA. Therefore, the computer-scoring procedures seemed only to reduce the mean IQ scores and not to affect the psychometric properties of the test.

Further evidence concerning the validity of the computer answer analysis algorithms can be obtained by viewing the effectiveness of these scoring procedures for correctly determining the adequacy of the input answers. A computer may score an answer as correct or incorrect from an examiner's viewpoint. The decision outcomes, in other words, may be a valid positive (the joint occurrence of correct classifications of an answer by the computer and the examiner), a false negative (where the computer classifies an answer as incorrect whereas in reality it is correct as judged by an examiner), a valid negative (where both the computer and the examiner score the answer as incorrect), or a false positive (where the computer scores an answer as correct when the examiner scores the answer as incorrect).

Table 20 presents a comparison of the effectiveness of the scoring system used for the present study with data from the initial field trial of the CB-SIT. The current CB-SIT correctly scored 91.6% of all the responses as either valid positives or

valid negatives. Of the responses, 6.7% of the responses were incorrectly scored as false negatives, whereas only 1.7% of these responses were misclassified as false positives. The item dictionary revisions based upon the field trial data showed an increase of 5% effectiveness over the initial field trial results. In addition, the number of false negatives was reduced from 12.3% to 6.7% in the present study. The percentage of false positives remained essentially the same on both occasions.

TABLE 20
Comparison of Computer and Examiner Scoring Methods for Initial Field Trial and Present Study

Item Type	Classification	Present Study	Field Trial
All items	Valid positive or negative	91.6%	87.2%
	False negatives	6.7%	12.3%
	False positives	1.7%	.5%
Verbal items	Valid positive or negative	88.5%	83.8%
	False negatives	9.4%	15.2%
	False positives	2.1%	1%
Numerical items	Valid positive or negative	96.4%	90.6%
	False negative	2.5%	9.4%
	False positive	1.1%	---

Based upon this scoring algorithm analysis, certain modifications are being implemented with the CB-SIT program. For items requiring word/phrase input, the item dictionaries are being enlarged to encompass a larger range of student answers. Additional keywords and phrases are being added to the storage buffers in an attempt to reduce the number of false negative classifications by the computer.

The majority of scoring errors for the numerical input items was due to the student inability to enter these answers in the manner specified in the program operation instructions. Therefore, the answer analysis routines would inappropriately score an answer as incorrect. Student direction revisions are being implemented to more adequately instruct students on the appropriate input for their numerical answers.

Included within the revised CB-SIT presentation strategy is the inclusion of a scoring system to determine the student's IQ score immediately as the student completes the test. Preliminary implementation activities are currently being performed on the designed scoring algorithms.

CHAPTER V
DISCUSSION

The present study focused on the development and evaluation of a computer-based administration and scoring program for the Slosson Intelligence Test. Although the feasibility of such a technological approach has previously been documented for the automated administration of the WAIS (Elwood, 1969), the combination of computer administration and real-time response evaluation had not yet been documented. The significant feature of the approach reported in this investigation was that the diagnostic evaluation occurred within a terminal-oriented interaction between the student and the computer system.

The feasibility of a computer approach can be seen from the relationships between the CB-SIT and SIT scores for the same students. The magnitude of these correlations was in the order of $r = .75$ ($p < .01$) for the entire sample, as well as for each of the three experimental groups. It should be noted that these correlations were somewhat lower than reported by Orr (1969) for the automated administration of the WAIS in comparison to a traditional administration. However, Arnold, Steward, and Rawson (1969) found lower correlational relationships for the automated WAIS when dealing with a relatively homogeneous sample of college students. A similar restriction in range may also have accounted for the lowered relationships found in the present investigation. When the correction is made for the restricted range of individual differences, the correlation becomes .96 between the CB-SIT and the SIT.

With respect to the intelligence scores, since the intertest interval was about one week on the average, some memory or practice effects should be expected. This expectation was confirmed in finding a significant increase in the scores over the three test sessions. There was no apparent effect of testing with respect to the WAIS as can be noted by the equivalent mean WAIS FS-IQ's in all groups. However, test-retest effects were clearly indicated in all three groups for the different administrations of the SIT. When the CB-SIT preceded the SIT (i.e., Groups 1 and 2), a mean increase of approximately 10 points was noted for the SIT scores. In contrast, when the SIT preceded the CB-SIT (Group 3) a mean decrease of approximately seven points was found.

The observed increase in Groups 1 and 2 may partially have been a function of the stimulus characteristics of the computer item presentation. Items were visually presented on the cathode ray tube, whereas, in a traditional SIT administration, item presentation was verbal in nature. Furthermore, the amount of presentation time was longer for the CB-SIT. The items remained on the screen until the student responded to the item. Thus, the nature of the computer visual presentation time may have reduced the memory requirements for each item, and therefore, maximized test-rest practice effects. The unexpected decrease in Group 3 may have been a function of both the lack of the visual memory effects as well as the increased difficulty of the computer testing situation.

Other researchers in the automated testing field have also noted the increased difficulty and presentation order effects of automated testing procedures. For example, Orr (1969) found a larger mean IQ change or test-retest effect for Ss who were first

tested with the automated WAIS and then later tested with a traditional WAIS examination. This was especially true in the Performance subtests of the WAIS in which an average of over 12 IQ points was gained in being tested with the automated WAIS first. The greater difficulty of the automated Performance section tended to inflate the apparent gain in the later examiner administration of the WAIS. In addition, Elwood (1969) found that WAIS performance scores were lower in the automated testing booth programmed for WAIS presentation in comparison to scores obtained in a traditional examiner administration of the WAIS.

In terms of concurrent validity indices with the WAIS, equivalent relationships were shown for the CB-SIT and SIT with WAIS VIQ's, PIQ's, and FS-IQ's. The magnitude of the differences between the correlations of the CB-SIT and SIT with the WAIS were not shown to be statistically significant. This was true for the entire sample as well as for the three experimental groups. Thus, it would appear that a standardized individual test of intelligence can be computer administered with no apparent loss in concurrent validity.

Further evidence concerning the validity of computer techniques to the assessment of intelligence can be seen in the predictive validity data. Although the equations with CB-SIT and SIT scores did not differ significantly from each other, the best predictor of college grade point average in terms of R^2 was the CB-SIT. Moreover, the WAIS IQ scores were not shown to contribute predictive variance over and above that of the sex factor. The extreme restriction in range on the WAIS scores ($\bar{X}$ = 120.7, **sd** = 6.7) may account for this lack of predictive validity. There was a somewhat larger range for the CB-SIT and SIT scores. The CB-SIT was shown to have a mean of 121.2 with a standard deviation of 8.6. In comparison, the SIT scores showed a mean of 131.6 with a standard deviation of 9.7.

One significant feature of the CB-SIT was the provision for the real-time analysis of the students' answers. A major question in this regard concerns the adequacy of the present word/phrase dictionaries for scoring and deriving valid IQ scores for the CB-SIT. Analysis of the examiner-derived IQ scores for the CB-SIT yielded a correlation of .89 for the entire sample of 48 college students. This is to be contrasted with the initial field trial results which showed a correlation of .83 between the examiner-derived and computer-derived scores. The mean computer-derived CB-SIT score was 121.2 (**sd** = 8.6). In comparison, the mean examiner-derived score for the CB-SIT was 125.7 (**sd** = 9.6). The 11 point difference observed in the initial field trial between the two scoring procedures was reduced to 4.5 points in the present study.

The scoring procedures for the CB-SIT still tend to produce somewhat lower estimates of the IQ score than SIT examiners. However, when the data was reanalyzed using the examiner-derived scores for the CB-SIT, equivalent results were found for the evaluation of test administration procedures, for the concurrent validity evaluation in terms of college grade point average. Thus, the feasibility and validity of the automated scoring program are clearly demonstrated for this college population.

Another purpose of the present study was to examine certain affective variables within the computer and examiner testing situations. Previous writers have conjectured about the objective and confidential aspects of the computer interactive experience. The data from the present investigation do not support these contentions. The significant

decline of the state anxiety scores over the three testing sessions is in agreement with the data reported in Table 2 in which a significant increase in the IQ scores was noted over the same sessions. Thus, as the tasks became less difficult, level of state anxiety decreased. However, the computer testing experience lead to higher levels of state anxiety in comparison to either of the two examiner administered testing sessions. This was true for the pre- and post-measurement periods indicating that this particular computer testing experience was perceived by the students as somewhat stressful in nature.

The student attitude results were fairly consistent with these anxiety results. Generally, attitudes toward examiner testing increased from the pre- to the post-measurement periods, whereas attitudes toward the computer testing procedures decreased over the same measurement periods. However, the triple interaction of tests x test sessions x measurement periods indicated that the student attitudes were differentially influenced by the experimental procedures. In addition, the results also point out the fact that the attitudinal reactions were also different between the two examiner administered tests. Anecdotal student comments revealed that they generally liked the examiner tests better, and they expressed a marked preference for the WAIS because of the varied nature of the test materials.

The somewhat negative affective reactions toward the computer testing procedures in terms of levels of state anxiety and student attitudes may partly have been a function of the particular programming methodology used in the present CB-SIT. Factors such as the increased presentation time within the CB-SIT as compared to the SIT (50 minutes versus 25 minutes), the nature of some of the computer feedback statements, the amount of time allowed for the student answers, and unfamiliarity with the terminal typewriter may have caused certain of these adverse reactions. Other adverse reactions seem to have been precipitated by the lack of feedback within the computer testing experience.

These data also point to another interpretation concerning the affective nature of computer examination techniques. The computer may not be perceived as being less threatening, more confidential, etc. for all testing situations. The confidential perception of the computer may be especially true for certain personality type tests in which a number of very personal questions are being elicited, but may not be true for the computer presentation of cognitive ability tests. Although logically it would seem that the failure aspects of the testing session would be reduced in a computer environment, it could be that the students know that they will be eventually evaluated by a psychologist and that the confidentiality of their data are not restricted to the computer. Obviously, this test-dependent interpretation concerning the affective nature of the computer evaluation procedures is an empirical question that requires further investigation.

Major CB-SIT programming presentation revisions are currently being implemented based upon the data from the present investigation. The new presentation strategy will incorporate basal and ceiling cut-off strategies as well as item revisions. These changes should substantially reduce the time factor of test presentation. In addition, revised instructions and procedures are being developed to more adequately familiarize students with terminal and program characteristics and operations. The

revised CB-SIT will then be compared with the current data from the CB-SIT in terms of the obtained affective anxiety and attitude reactions.

One exploratory aspect on the present study was to determine the relationships of examiner ratings of test behavior to observed performance and to the student attitude and anxiety reactions to the testing sessions. As to be expected, the examiner ratings were significantly related to the performance of the subjects. However, the ratings were negatively related to the sex of the student. Female students were rated higher than the male students by the WAIS examiners. Sex of the student was also negatively related, but not significantly with the SIT examiner ratings.

In terms of the affective measures, the examiner ratings were shown to be positively related to the student attitude measures. Students who positively rated the testing experience were also rated more favorably by their particular examiners. In addition, the ratings were negatively related to the state anxiety scores. Higher ratings were given for students who reported lower anxiety levels.

These results also lend credence to the quality of the testing experience that determines the nature of affective reactions. Performance of the student was not shown to correlate with their reported affective attitude and anxiety reactions, except for a significant negative correlation ($r = -.29$) between WAIS FS-IQ's and post state anxiety. These data would seem to indicate that the affective reactions were not essentially performance-based in the present investigation, and that affective reactions may be test and context procedurally specific. Sattler and Theye (1967) and Masling (1960) have also pointed out that the effects of examiner-student interactions must be interpreted in light of the particular type of test and the ability that is being measured. Therefore, in computer-based evaluation, it would also seem that there might be a complex interaction between the nature of the test and the quality of the particular programming procedures that, in turn, determines the affective nature of the computer experience.

The results of the present investigation demonstrate the feasibility and validity of an interactive computer approach to the assessment of intelligence. However, the generalizability of these computer procedures is subject to further investigation. A real-time response analysis program demands a fairly sophisticated performance in terms of input. The word/phrase dictionary matching approach to the scoring of the input words, phrases, sentences, demands that the input performance level be fairly good in terms of spelling. This skill demand may only be appropriate for high level ability students. Age of the student would probably be an important factor in this respect.

Another important issue raised by the present investigation is that the major research in the automated testing area has been concerned with the simulation of standard psychometric instruments. Although the results of this study and others in the area support the reliability and validity of this approach, it does raise some important questions. For instance, automation of a test does seem to change the test sufficiently to demand a renorming procedure for the automated test.

Given the wide controversy over the uses of computers in psychological evaluation, research should focus on a determination of the types of abilities that can be adequately assessed by computer techniques and which can only be assessed by the examiner. Another question would relate to the types of students that can be efficiently and appropriately examined by the computer. In addition, the amount of

examiner-student- computer interaction to produce the most efficient and effective indices should be investigated to insure that the students and examiners place the appropriate confidence in the computer.

CHAPTER VI
SUMMARY

The general purpose of the present study was to investigate a computer application to the area of individualized intelligence testing. Primary emphasis was placed on questions relating to the feasibility, reliability, and validity of a computer-based administration and scoring program for the Slosson Intelligence Test (Slosson, 1963). Specifically, the evaluation focused on concurrent validity relationships with the Wechsler Adult Intelligence Scale (Wechsler, 1955). In addition, predictive validity in terms of college grade point average was compared in terms of forecasting efficiency. Certain affective variables (student attitudes and anxiety) were studied to determine their relationship to the computer-based and traditional examiner test administration procedures. Examiner ratings of student testing behaviors were related to the attitudes and anxiety variables.

The computer-based Slosson Intelligence Test (CB-SIT) was designed to operate with an IBM 1500 Instructional System (IBM, 1967). The test items were individually presented on a cathode ray tube. Students entered their answers for immediate computer evaluation. The answer analysis algorithms were based upon a key word/phrase dictionary for each item which was developed from previous test item protocols. If the student was scored as either correct or incorrect, the program continued to the next item. If the answer was scored as partially correct, the student was instructed by the computer to more fully explain his response.

Forty-eight undergraduate students (24 males and 24 females) were individually tested with the WAIS, SIT, and CB-SIT. A latin square design was employed to counterbalance potential order and sequence effects of the three different intelligence test administrations. In addition, each testing session included an administration of an attitude scale and anxiety scale both before and after each testing period. A time period of approximately one week separated the different test administrations.

Within the context of the latin square design, the CB-SIT was shown to correlate .75 with a traditional examiner administration of the SIT. More importantly, equivalent concurrent validity relationships with the WAIS were shown as follows: the CB-SIT correlated .54 with the WAIS and the SIT correlated .52 with the WAIS with a college population. In terms of predictive validity, the best predictor of college grade point average in terms of R^2 was the CB-SIT.

In terms of the affective measures, the computer testing experience lead to higher levels of state anxiety in comparison to either of the two examiner testing sessions. The student attitude results were fairly consistent with the anxiety results. Generally, attitudes toward examiner testing increased for the pre-to-post measurement periods, whereas attitudes toward the computer testing procedures decreased over the same measurement periods.

Examiner ratings of test behavior were found to be significantly related to the performance of the Ss. In addition, these ratings were also shown to be positively related to the student attitude measures and negatively related to the state anxiety scores.

The results of the present investigation demonstrated the feasibility and validity of an interactive computer approach to the assessment of intelligence. In addition, the attitudinal and anxiety data indicated that there might be a complex interaction between the nature of the test and the quality of the computer programming procedures that in turn determines the affective nature of automated testing procedures.

REFERENCES

Archambault, F. X. A computerized approach to scoring verbal responses to the *Torrance tests of creative thinking*. Paper presented at the meeting of the American Educational Research Association, Minneapolis, 1970.

Arnhoff, F. N. Reassessment of the trilogy: Need, supply and demand. *American Psychologist*, 1968, **23**, 312-316.

Arnold, D. J., Stewart, L. T., & Rawson, H. E. Some exploratory research on the functional usages of the automated testing booth. Paper presented at the meeting of the Indiana Psychological Association, Indianapolis, April 1969.

Boneau, A. Psychology's manpower: Report on the 1966 national register of scientific and technical personnel. *American Psychologist*, 1968, **23**, 325-334. (a)

Boneau, A. The educational base: supply for the demand. *American Psychologist*, 1968, **23**, 308-311. (b)

Cogswell, J. F., & Estavan, D. P. Explorations in computer-assisted counseling. System Development Corporation, TM-2582, 1965.

Cronbach, L. J. *Essentials of psychological testing*. New York: Harper & Row, 1960.

DeCecco, J. P. *The psychology of learning and instruction: Educational psychology*. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

DeLapa, G. The Slosson Intelligence Test: A screening and retesting technique for slow learners. *Journal of School Psychology*, 1968, **6** (3), 224-225.

Eber, H. W. Computer reporting of 16 PF data. Paper presented at the meeting of the American Psychological Association, Los Angeles, January, 1964.

Elwood, D. L. Automation of psychological testing. *American Psychologist*, 1969, **24** (3), 287-289.

Finney, J. C. Methodological problems in programmed composition of psychological test reports. *Behavioral Science*, 1967, **12**, 142-152.

Fowler, R. D., Jr. Computer processing and reporting of personality test data. Paper presented at the meeting of the American Psychological Association, Los Angeles, January 1964.

Fowler, R. D., Jr. The current status of computer interpretation of psychological tests. *American Journal of Psychiatry*, 1969, **125** (7), 21-27.

Gallagher, P. D. An investigation of instructional treatments and learner characteristics in a computer-managed instruction course. *Technical Report No. 12*, Computer-Assisted Instruction Center, Florida State University, 1970.

Goldberg, J. B. Computer analysis of sentence completions. *Journal of Projective Techniques and Personality Assessment*, 1966, **30**, 37-45.

Gorham, D. R. Validity and reliability studies of a computer-based scoring system for inkblot responses. *Journal of Consulting Psychology*, 1967, **31**, 65-70.

Griffith, H. R., & Elwood, D. L. The automated administration of the Wechsler Adult Intelligence Scale. Paper presented at the meeting of the Indiana Psychological Association, Indianapolis, April, 1969.

Guertin, W. H., Frank, G. H., & Rabin, A. I. Research with the Wechsler-Bellevue Intelligence Scale: 1950-1955. *Psychological Bulletin*, 1962, **59** (1), 1-26.

Guertin, W. H., Ladd, C. E., Frank, G. H., Rabin, A. I., & Heister, D. C. Research with the Wechsler Intelligence Scale for Adults:: 1960-1965. *Psychological Bulletin*, 1966, **66** (5), 385-409.

Guertin, W. H., Rabin, A. I., Frank, G. H., & Ladd, C. E. Research with the Wechsler Intelligence Scales for Adults:: 1955-1960. *Psychological Bulletin*, 1962, **59** (1), 1-26.

Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.

Hammil, D. The Slosson Intelligence Test as a quick estimate of mental ability. *Journal of School Psychology*, 1967, **7** (4), 171-175.

Holt, R. R. Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology*, 1958, **56**, 1-12.

Holtzman, W. H. Can the computer supplant the clinician? *Journal of Clinical Psychology*, 1960, **16**, 119-122.

International Business Machines Corporation. *IBM 1500 Coursewriter II Author's Guide*. New York: IBM, 1967.

Kaufman, H., & Ivanoff, J. The Slosson Intelligence Test as a screening instrument with a rehabilitation population. *Exceptional Children*, 1969, **35** (9), 35-41.

Lawler, R. M. Computer assessment of human sexuality knowledge. Unpublished manuscript, Florida State University, 1968.

Lawler, R. M., O'Neil, H. F., Jr., & Hansen, D. N. Anxiety and confidence in computer-assisted learning. Paper presented at the meeting of the American Educational Research Association, Minneapolis, March 1970.

Leherissey, B., O'Neil, H. F., Jr., & Hansen, D. N. Effects of memory support upon anxiety and performance in computer-assisted learning. Paper presented at the meeting of the American Educational Research Association, Minneapolis, March 1970.

Mackinnon, R. C., & Elliott, C. A. A comparison of an achievement battery with two tests of ability with educable mental retardates. Unpublished manuscript, Florida State University, 1969.

Marks, P. A., & Seeman, W. *The actuarial description of abnormal personality*. Baltimore: Williams & Williams, 1963.

Masling, J. M. The influence of situational and interpersonal variables in projective testing. *Psychological Bulletin*, 1960, **57**, 65-85.

Mathis, A., Smith, T. A., & Hansen, D. N. College students' attitudes toward computer-assisted instruction. *Journal of Educational Psychology*, 1970, **61** (1), 46-51.

McArthur, C. Clinical versus actuarial prediction. In Proceedings, 1955 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1956. Pp. 99-106.

McNemar, Q. *Psychological statistics*. (3rd ed.) New York: John Wiley & Sons, Inc., 1962.

Meehl, P. Wanted - a good cookbook. *American Psychologist*, 1956, **11**, 263-272.

O'Keefe, S. L. An inquiry into the use of the Slosson Intelligence Test with fast learners. Unpublished master's thesis, Ohio State University, 1967.

O'Neil, H. F., Jr. Effects of stress on state anxiety and performance in computer-assisted learning. *Technical Report No. 6*, Computer-Assisted Instruction Center, Florida State University, 1970.

O'Neil, H. F., Jr., Hansen, D. N., & Spielberger, C. D. The effects of state and trait anxiety on computer-assisted learning. Paper presented at the meeting of the American Educational Research Association, Los Angeles, 1969.

O'Neil, H. F., Jr., Spielberger, C. D., & Hansen, D. N. The effects of state-anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology*, 1969, **60**, 343-350.

Orr, T. B. A comparison of the automated method and the face-to-face method of administering the Wechsler Adult Intelligence Scale. Paper presented at the meeting of the Indiana Psychological Association, Indianapolis, April 1969.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning* Urbana: University of Illinois Press, 1957.

Peck, R. F., & Veldman, D. J. An approach to psychological assessment by computer. *Research Memorandum No. 10*, Mental Health in Teacher Training Project. University of Texas, 1961.

Piotrowski, Z. A. Digital computer interpretation of inkblot test data. *Psychiatric Quarterly*, 1964, **38**, 1-26.

Roget, P. (Ed.). *Roget's International Thesaurus.* New York: Crowell, 1962.

Rome, H. P., Swenson, W. M., Mayaya, P., McCarthy, C. E., Pearson, J. S., & Keating, R. F. Symposium on automation techniques in personality assessment. *Proceedings of the Mayo Clinic*, 1962, **37**, 61-82.

Sattler, J. M., & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. *Psychological Bulletin*, 1967, **68** (5), 347-360.

Slosson, R. L. *The Slosson Intelligence Test for Children and Adults*. East Aurora: Slosson Education Publication, 1963.

Smith, M. S. The computer and the TAT. *Journal of School Psychology*, 1968, **6** (3), 206-214.

Smith, R. E. Examination by computer. *Behavioral Science*, 1963, **8**, 76-79.

Soule, R. *Soule's Dictionary of English Synonyms*. Boston: Little, Brown, & Co., 1966.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. *The state-trait anxiety inventory (Test manual)*. Palo Alto, California: Consulting Psychologists Press, 1970.

Spielberger, C. D., Lushene, R. E., & McAdoo, W. G. Theory and measurement of anxiety states. In R. B. Cattell (Ed.), *Handbook of Modern Personality Theory*. Chicago: Aldine, In press, 1971.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. *The general inquirer: A computer approach to content analysis*. Cambridge, Mass.: MIT Press, 1966.

Terman, L. M., & Merrill, M. A. *Revised Stanford-Binet Intelligence Scale, third revision.* Boston: Houghton Mifflin, 1960.

Veldman, D. J. Computer-based sentence completion interviews. *Journal of Counseling Psychology*, 1967, **14** (2), 153-157.

Vernon, P. E. Education and the psychology of individual differences. *Harvard Educational Review*, 1958, **28**, 91-104.

Wechsler, D. *Wechsler Adult Intelligence Scale manual.* New York: Psychological Corporation, 1955.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

APPENDICES

APPENDIX A
SHORT FORM OF THE STATE-TRAIT ANXIETY INVENTORY

APPENDIX A
PRETEST
Self-Analysis Questionnaire
Form X-1

NAME _____ DATE _____

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you feel right now, that is, at this moment.

There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

	Not at all	Somewhat	Moderately so	Very much so
1. I am relaxed	1	2	3	4
2. I am tense	1	2	3	4
3. I am jittery	1	2	3	4
4. I feel calm	1	2	3	4
5. I feel at ease	1	2	3	4

APPENDIX A
POSTTEST
Self-Analysis Questionnaire
Form X-1

NAME_____ _____ _____DATE_____

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you felt during the test which you just completed.

There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your feelings during the testing situation.

	Very much so	Moderately so	Somewhat	Not at all
1. I was relaxed	1	2	3	4
2. I was tense	1	2	3	4
3. I was jittery	1	2	3	4
4. I felt calm	1	2	3	4
5. I felt at ease	1	2	3	4

APPENDIX B
STATE-TRAIT ANXIETY INVENTORY A-TRAIT SCALE

APPENDIX B
TRAIT ANXIETY INVENTORY
Form X-2

© Copyright, 1970, by
Consulting Psychologists Press
Inc. Permission granted, 1971.

CAI Number_____ Date_____

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement to indicate how you generally feel.

There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe how you generally feel.

	Almost Never	Sometimes	Often	Almost Always
1. I feel pleasant	1	2	3	4
2. I tire quickly	1	2	3	4
3. I feel like crying	1	2	3	4
4. I wish I could be as happy as others seem to be	1	2	3	4
5. I am losing out on things because I can't make up my mind soon enough	1	2	3	4
6. I feel rested	1	2	3	4
7. I am "calm, cool, and collected"	1	2	3	4
8. I feel that difficulties are piling up so that I cannot overcome them	1	2	3	4
9. I worry too much over something that really doesn't matter	1	2	3	4
10. I am happy	1	2	3	4
11. I am inclined to take things hard	1	2	3	4
12. I lack self-confidence	1	2	3	4

	Almost Never	Sometimes	Often	Almost Always
13. I feel secure	1	2	3	4
14. ' try to avoid facing a crisis or difficulty	1	2	3	4
15. I feel blue	1	2	3	4
16. I am content	1	2	3	4
17. Some unimportant thought runs through my mind and bothers me	1	2	3	4
18. I take disappointments so keenly that I can't put them out of my mind	1	2	3	4
19. I am a steady person	1	2	3	4
20. I become tense and upset when I think about my recent concerns and interests	1	2	3	4

APPENDIX C
ATTITUDE SCALE

APPENDIX C
ATTITUDE SCALE

Name _____

Student Number _____

Date _____

INSTRUCTIONS

The purpose of this study is to measure the meanings of certain things to various people by having them judge them against a series of descriptive scales. In taking this test, please make your judgments on the basis of what these things mean to you. On each page you will find a different concept to be judged and beneath it a set of scales. You are to rate the concept on each of these scales in order.

Here is how you are to use these scales:

If you feel that the concept at the top of the page is very closely related to one end of the scale, you should place your check-mark as follows:

fair ___X___ _____ _____ _____ _____ _____ _____unfair

or

fair _____ _____ _____ _____ _____ _____ ___X___unfair

If you feel that the concept is quite closely related to one or the other end of the scale (but not extremely) you should place your check-mark as follows:

strong_____ ___X___ _____ _____ _____ _____ _____weak

or

strong_____ _____ _____ _____ _____ ___X___ _____weak

If the concept seems only slightly related to one side as opposed to the other side (but is not really neutral), then you should check as follows:

active_____ _____ ___X___ _____ _____ _____ _____passive

or

active_____ _____ _____ ___X___ _____ _____ _____passive

The direction toward which you check, of course, depends upon which of the two ends of the scale seem most characteristic of the thing you're judging.

If you consider the concept to be neutral on the scale, both sides of the scale equally associated with the concept, or if the scale is completely irrelevant, unrelated to the concept, then you should place your check-mark in the middle space:

safe _____ _____ _____ __X__ _____ _____ _____dangerous

IMPORTANT: (1) Place your check-mark in the middle of spaces, not on the boundaries:

	THIS		NOT THIS			
_____	__X__	_____	_____X_____	_____	_____	

(2) Be sure you check every scale for every concept—do not omit any!
(3) Never put more than one check-mark on a single scale.

Sometimes you may feel as though you've had the same item before on the test. This will not be the case, so do not look back and forth through the items. Do not try to remember how you checked similar items earlier in the test. Make each item a separate and independent judgment. Work at fairly high speed through this test. Do not worry or puzzle over individual items. It is your first impression, the immediate "feelings" about the items, that we want. On the other hand, please do not be careless, because we want your true impressions.

COMPUTER TESTING

	extremely:	quite:	slightly:	neutral:	slightly:	quite:	extremely:	
FAST	_____	____	_____	_____	_____	____	_____	SLOW
DULL	_____	____	_____	_____	_____	____	_____	INTERESTING
TENSE	_____	____	_____	_____	_____	____	_____	RELAXED
GOOD	_____	____	_____	_____	_____	____	_____	BAD
UNFAIR	_____	____	_____	_____	_____	____	_____	FAIR
DEEP	_____	____	_____	_____	_____	____	_____	SHALLOW
VALUABLE	_____	____	_____	_____	_____	____	_____	WORTHLESS
ACTIVE	_____	____	_____	_____	_____	____	_____	PASSIVE
EASY	_____	____	_____	_____	_____	____	_____	DIFFICULT

	extremely:	quite:	slightly:	neutral:	slightly:	quite:	extremely:	
INFLEXIBLE	_____	____	_____	_____	_____	____	_____	FLEXIBLE
EXCITING	_____	____	_____	_____	_____	____	_____	BORING
UNPLEASANT	_____	____	_____	_____	_____	____	_____	PLEASANT
DEPRESSING	_____	____	_____	_____	_____	____	____	ENCOURAGING
SAFE	_____	____	_____	_____	_____	____	_____	DANGEROUS
WEAK	_____	____	_____	_____	_____	____	_____	STRONG

APPENDIX D
EXAMINER RATING SCALE

APPENDIX D
EXAMINER RATING SCALE

	extremely 1	quite 2	slightly 3	neutral 4	slightly 5	quite 6	extremely 7
1.	Absorbed by task				_____		Easily distracted
2.	Initiates activity				_____		Waits to be told
3.	Quick to respond				_____		Urging needed
4.	Socially confident				_____		Shy, reserved, reticent
5.	Realistically self-confident				_____		Overconfident
6.	Comfortable in adult company				_____		Ill-at-ease
7.	Assured				_____		Anxious about success
8.	Persistent				_____		Gives up easily
9.	Reacts to failure realistically				_____		Withdrawing, hostile
10.	Eager to continue				_____		Seeks to terminate
11.	Challenged by hard task				_____		Prefers only easy tasks
12.	Needs minimum of commendation				_____		Needs constant praise and encouragement

VITA

BIOGRAPHICAL DATA

NAME:	John Joseph Hedl, Jr.
DATE OF BIRTH:	August 11, 1945
PLACE OF BIRTH:	Upper Darby, Pennsylvania
MARITAL STATUS:	Single
SOCIAL SECURITY NO.:	159-36-7276
ADDRESS:	Computer-Assisted Instruction Center Tully Building, Room 1-A Florida State University Tallahassee, Florida 32306 Phone No. (904) 599-3660

ACADEMIC HISTORY

B.S., Psychology, The Pennsylvania State University, 1967

M.S., School Psychology, Florida State University, 1969

Ph.D., School Psychology, Florida State University, June, 1971

PROFESSIONAL EXPERIENCE

NIMH Graduate Fellow in School Psychology, Department of Psychology, Florida State University, 1967-68. Training in School of Psychological Services with an emphasis on behavior modification.

Research Associate, Computer-Aided Instruction Laboratory, Harvard University, Summer, 1968, under the direction of Laurence M. Stolurow. Worked extensively with the automation of diagnostic reading tests. Served as an instructor to Philadelphia and Boston teachers in a summer institute concerning the programming and educational uses of CAI.

U.S. Office of Education Fellow in Computer-Assisted Instruction, Florida State University, 1968-69. Extensive training in educational technology including author programming, data analysis procedures, and design, implementation, and evaluation procedures for CAI and CMI.

NIMH Graduate Fellow in School Psychology, Department of Psychology, Florida State University, 1969 to present. Conducted an evaluation of a token reinforcement system operating in a Tallahassee elementary school. Worked extensively with the automation of an individualized intelligence test. Participated in the preparation of the following two research proposals: "The Development

of a Computer-Based Intelligence Test"; and "The Effects of Anxiety Reduction Techniques on State Anxiety and Performance in Computer-Assisted Learning and Evaluation" submitted to the United States Office of Education.

PROFESSIONAL AFFILIATIONS

American Educational Research Association (AERA)

PRESENT RESEARCH INTERESTS

Computer Assessment of Intelligence and Personality; Educational Technology including CAI and CMI; School Psychological Services.

PUBLICATIONS

Hansen, D. N., Hedl, J. J., Jr., & O'Neil, H. F., Jr. Computer processes in education. *Computers and Automation*, 1971, March, in press.

Hedl, J. J., Jr., A systematic investigation of three facets of programmed instruction: Tutorial assistance of study, explanation of incorrect answers, and the spacing of high-difficulty frames. *Technical Report No. 3*, CAI Center, Florida State University, January, 1970.

Hedl, J. J., Jr., Review of Tomkins, S. S., & Messick, S. (Eds.), *Computer Simulation of Personality*. New York: John Wiley & Sons, Inc., 1963. *In Journal of School Psychology*, 1968, **6**(3), 220-221.

PAPERS PRESENTED AT PROFESSIONAL MEETINGS

Hedl, J. J., Jr., O'Neil, H. F., Jr., & Hansen, D. N. Computer-based intelligence testing. Paper presented at the annual meeting of the American Educational Research Association, February, 1971.

Hansen, D. N., Hedl, J. J., Jr., & O'Neil, H. F., Jr. Review of automated testing. Paper presented at the annual meeting of the American Educational Research Association, February, 1971.

Hedl, J. J., Jr., O'Neil, H. F., Jr., & Hansen, D. N. The affective nature of computer-based intelligence testing procedures. Paper accepted for presentation to the American Psychological Association Convention, Washington, D.C., 1971.

DISTRIBUTION LIST

NAVY

4 Director, Personnel and Training Research Programs
Office of Naval Research
Arlington, VA 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Director
ONR Branch Office
536 South Clark Street
Chicago, IL 60605

1 Commander
Operational Test and Evaluation Force
U.S. Naval Base
Norfolk, VA 23511

1 Commanding Officer
Naval Medical Neuropsychiatric Research Unit
San Diego, CA 92152

1 Chief of Naval Operations (Op-98)
Department of the Navy
Washington, D.C. 20350
ATTN: Dr. J. J. Collins

1 Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, Maryland 20014

1 Chief of Naval Training
Naval Air Station
Pensacola, Florida 32508
ATTN: Capt. Allen E. McMichael

1 Mr. S. Friedman
Special Assistant for Research & Studies
OASN (M&RA)
The Pentagon, Room 4E794
Washington, D.C. 20350

6 Director
Naval Research Laboratory
Washington, D.C. 20390
ATTN: Library, Code 2029 (ONRL)

6 Director
Naval Research Laboratory
Washington, D.C. 20390
ATTN: Technical Information Division

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, VA 22314

1 Behavioral Sciences Department
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

1 Chief
Bureau of Medicine and Surgery
Code 513
Washington, D.C. 20390

1 Director
Education and Training Sciences Department
Naval Medical Research Institute
National Naval Medical Center
Building 142
Bethesda, MD 20014

2 Technical Director
Personnel Research Division
Bureau of Naval Personnel
Washington, D.C. 20370

2 Technical Library (Pers-11B)
Bureau of Naval Personnel
Department of the Navy
Washington, D. C. 20360

1 Technical Director
Naval Personnel Research and Development Laboratory
Washington Navy Yard, Building 200
Washington, D.C. 20390

1 Chief, Naval Air Reserve Training
Naval Air Station
Box 1
Glenview, IL 60026

1 Chief
Naval Air Technical Training
Naval Air Station
Memphis, TN 38115

1 Commander, Naval Air Systems Command
Navy Department, AIR-413C
Washington, D.C. 20360

1 Commanding Officer
Naval Air Technical Training Center
Jacksonville, FL 32213

1 Chief of Naval Air Training
Code 017
Naval Air Station
Pensacola, FL 32508

1 Research Director, Code 06
Research and Evaluation Department
U.S. Naval Examining Center
Building 2711 – Green Bay Area
Great Lakes, IL 60088
ATTN: C. S. Winiewicz

1 Commander
Submarine Development Group Two
Fleet Post Office
New York, NY 09501

1 Technical Library
Naval Ship Systems Command
National Center, Building 3 Room 3 S-08
Washington, D. C. 20360

1 Col. James Marsh, USMC
Headquarters Marine Corps (A01M)
Washington, D.C. 20380

1 Dr. James J. Regan, Code 55
Naval Training Device Center
Orlando, FL 32813

3 Commanding Officer
Naval Personnel and Training Research Laboratory
San Diego, CA 92152

1 Chairman
Behavioral Science Department
Naval Command and Management Div.
U.S. Naval Academy
Luce Hall
Annapolis, MD 21402

1 Superintendent
Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

1 Information Systems Programs
Code 437
Office of Naval Research
Arlington, VA 22217

1 Commanding Officer
Service School Command
U.S. Naval Training Center
San Diego, CA 92133

1 Technical Library
Naval Ordnance Station
Indian Head, MD 20640

1 Mr. George N. Graine
Naval Ship Systems Command (SHIP 03H)
Department of the Navy
Washington, D.C. 20360

1 Col. George Caridakis
Director, Office of Manpower Utiliz.
Headquarters, Marine Corps (A01H)
MCB
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code AX)
Commandant of the Marine Corps
Washington, D.C. 20380

ARMY

1 Behavioral Sciences Division
Office of Chief of Research and Development
Department of the Army
Washington, D. C. 20310

1 U.S. Army Behavior and Systems Research Laboratory
Commonwealth Building, Room 239
1320 Wilson Boulevard
Arlington, VA 22209

1 Director
Behavioral Sciences Laboratory
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760

1 Armed Forces Staff College
Norfolk, VA 23511
ATTN: Library

1 Director of Research
US Army Armor Human Research Unit
ATTN: Library
Bldg. 2422 Morande Street
Fort Knox, KY 40121

1 Commandant
U.S. Army Adjutant General School
Fort Benjamin Harrison, IN 46216
ATTN: ATSAG-EA

1 Division of Neuropsychiatry
Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Washington, D. C. 20012

AIR FORCE

1 AFHRL (TR/Dr. G. A. Eckstrand)
Wright-Patterson Air Force Base
Ohio 45433

1 AFSOR (NL)
1400 Wilson Boulevard
Arlington, VA 22209

1 HQ, AFSC (SDEC)
Andrews Air Force Base
Washington, D.C. 20330

1 Personnel Research Division (AFHRL)
Lackland Air Force Base
San Antonio, TX 78236

1 Commandant
U.S. Air Force School of Aerospace Medicine
ATTN: Aeromedical Library
Brooks AFB, TX 78235

1 Director of Manpower Research
OASD (M&RA) (M&RU)
Room 3D960
The Pentagon
Washington, D.C.

1 AFHRL (TRT/Dr. Ross L. Morgan)
Wright-Patterson Air Force Base
Ohio 45433

1 Lt. Col. Robert R. Gerry, USAF
Chief, Instructional Technology Programs
Resources & Technology Division
(DPTBD DCS/P)
The Pentagon (Room 4C244)
Washington, D.C. 20330

1 Director
Air University Library (AUL-8110)
Maxwell Air Force Base,
Alabama 36112

1 Headquarters, Electronics Systems Division
ATTN: Dr. Sylvia Mayer/MCDS
L. G. Hanscom Field
Bedford, MA 01730

OTHER GOVERNMENT

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh Street, S.W.
Washington, D. C. 20591

1 Dr. Andrew R. Molnar
Computer Innovation in Education Section
Office of Computing Activities
National Science Foundation
Washington, D.C. 20550

1 Dr. Alvin E. Goins, Chief
Personality and Cognition Research Section
Behavioral Sciences Research Branch
National Institute of Mental Health
5454 Wisconsin Ave., Room 10A01
Washington, D.C. 20015

MISCELLANEOUS

1 Dr. John Annett
Department of Psychology
Hull University
Hull
Yorkshire, England

1 Dr. Bernard M. Bass
University of Rochester
Management Research Center
Rochester, NY 14627

1 Dr. Jaime Carboneil
Bolt, Bernanek and Newman
50 Moulton Street
Cambridge, MA 02138

1 ERIC Clearinghouse on
Educational Media and Technology
Stanford University
Stanford, CA 94305

1 Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
2001 "S" St. N.W.
Washington, D.C. 20037

1 Dr. Albert S. Glickman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Mr. Roy Ference
Room 2311
U.S. Civil Service Commission
Washington, D.C. 20415

1 Dr. Richard C. Atkinson
Department of Psychology
Stanford University
Stanford, California 94305

1 Dr. Mats Bjorkman
University of Umea
Department of Psychology
Umea 6, Sweden

1 Dr. David Weiss
University of Minnesota
Department of Psychology
Elliot Hall
Minneapolis, MN 55455

1 ERIC Clearinghouse on Vocational and Technical Education
The Ohio State University
1900 Kenny Road
Columbus, OH 43210
ATTN: Acquistion Specialist

1 Dr. Robert Glaser
Learning Research and Development Center
University of Pittsburgh 15213

1 Dr. Bert Green
Department of Psychology
Johns Hopkins University
Baltimore, MD 21218

1 Dr. Duncan N. Hansen
Center for Computer Assisted Instruction
Florida State University
Tallahassee, FL 32306

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
11428 Rockville Pike
Rockville, MD 20852

1 Human Resources Research Organization
Library
300 North Washington Street
Alexandria, VA 22314

1 Human Resources Research Organization
Division #4, Infantry
Post Office Box 2086
Fort Benning, Georgia 31905

1 Human Resources Research Organization
Division #6, Aviation (Library)
Post Office Box 428
Fort Rucker, Alabama 36360

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
Santa Barbara Research Park
6780 Cortona Drive
Goleta, CA 93017

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Len Rosenbaum
Psychology Department
Montgomery College
Rockville, MD 20850

1 Dr. Robert J. Seidel
Human Resources Research Organization
300 N. Washington Street
Alexandria, VA 22314

1 Dr. M. D. Havron
Human Sciences Research, Inc.
Westgate Industrial Park
7710 Old Springhouse Road
McLean, VA 22101

1 Human Resources Research Organization
Division #3
Post Office Box 5787
Presidio of Monterey, CA 93940

1 Human Resources Research Organization
Division #5, Air Defense
Post Office Box 6021
Fort Bliss, TX 77916

1 Dr. Roger A. Kaufman
Graduate School of Human Behavior
U.S. International University
8655 E. Pomerada Road
San Diego, CA 92124

1 Office of Computer Information
Center for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

1 Psychological Abstracts
American Psychological Association
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
University Park
Los Angeles, CA 90007

1 Dr. George E. Rowland
Rowland and Company, Inc.
Post Office Box 61
Haddonfield, NJ 08033

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, PA 19087

1 Benton J. Underwood
Department of Psychology
Northwestern University
Evanston, IL 60201

1 Mr. C. R. Vest
6225 Nelway Drive
McLean, Virginia 22101